DEX-0214

EJ

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/098339 A3

(51) International Patent Classification[7]: C12Q 1/68, G01N 33/574, C12N 15/11, C07K 14/47, C12N 15/85, C07K 16/30

(21) International Application Number: PCT/US01/19032

(22) International Filing Date: 12 June 2001 (12.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/602,877	22 June 2000 (22.06.2000)	US
09/687,507	12 October 2000 (12.10.2000)	US
09/778,381	6 February 2001 (06.02.2001)	US

(71) Applicant *(for all designated States except US)*: **CORIXA CORPORATION** [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **REED, Steven, G.** [US/US]; 2843 122nd Place N.E., Bellevue, WA 98005 (US). **XU, Jiangchun** [US/US]; 15805 S.E. 43rd Place, Bellevue, WA 98006 (US). **DILLON, Davin, C.** [US/US]; 18112 N.W. Montreux Drive, Issaquah, WA 98027 (US). **RETTER, Marc, W.** [US/US]; 33402 N.E. 43rd Place, Carnation, WA 98014 (US). **HARLOCKER, Susan, L.** [US/US]; 7522 13th Avenue W., Seattle, WA 98117 (US).

(74) Agents: **POTTER, Jane, E., R.**; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 31 July 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/098339 A3

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF BREAST CANCER

(57) Abstract: Compositions and methods for the therapy and diagnosis of cancer, particularly breast cancer, are disclosed. Illustrative compositions comprise one or more breast tumor polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cell that expresses such polypeptides, and T cells that are specific for cells expressing such polypeptides. The disclosed compositions are useful, for example, in the diagnosis, prevention and/or treatment of diseases, particularly breast cancer.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/19032

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68 G01N33/574 C12N15/11 C07K14/47 C12N15/85
C07K16/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q G01N C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 33869 A (CORIXA CORP) 8 July 1999 (1999-07-08) the whole document ---	1-9, 11-17
X	WO 98 18945 A (ABBOTT LAB) 7 May 1998 (1998-05-07) relating to SEQ ID NO.5 the whole document ---	1-8, 11-16
Y	WO 98 21331 A (AKERBLOM INGRID E ;HAWKINS PHILLIP R (US); INCYTE PHARMA INC (US);) 22 May 1998 (1998-05-22) abstract; claims 1-22 --- -/--	3-9, 11-17

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 December 2002	13. 03. 03

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bradbrook, D

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 98 33915 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); JI HONGJUN (US)) 6 August 1998 (1998-08-06) page 19, line 8 -page 29, line 26; claims 1-15	3-9, 11-17
A	DIATCHENKO L ET AL: "SUPPRESSION SUBTRACTIVE HYBRIDIZATION: A METHOD FOR GENERATING DIFFERENTIALLY REGULATED OR TISSUE-SPECIFIC CDNA PROBES AND LIBRARIES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 93, 1 June 1996 (1996-06-01), pages 6025-6030, XP002911922 ISSN: 0027-8424 page 6025, column 1, paragraph 1 -column 2, paragraph 2 page 6027, column 1, paragraph 3 -page 6028, column 1, paragraph 2; figure 1	1-9, 11-17
A	MARTIN K J ET AL: "A HYBRIDIZATION ARRAY ASSAY USING DIFFERENTIAL DISPLAY-IDENTIFIED MARKERS FOR EARLY DETECTION AND STAGING OF BREAST CANCER" PROCEEDINGS OF THE ANNUAL MEETING OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH, NEW YORK, NY, US, vol. 40, 10 April 1999 (1999-04-10), page 319 XP001026342 ISSN: 0197-016X abstract	1-9, 11-17
A	XU ET AL: "Identification of differentially expressed genes in human breast tumor using substraction and microarray" CHEMICAL ABSTRACTS + INDEXES, AMERICAN CHEMICAL SOCIETY. COLUMBUS, US, vol. 40, March 1999 (1999-03), page 319 XP002160220 ISSN: 0009-2258 abstract	1,2
E	WO 01 75171 A (MOLESH DAVID ALAN ;ZEHENTNER BARBARA (US); CORIXA CORP (US); DILLO) 11 October 2001 (2001-10-11) claims 1-46; examples 1-12	1-9, 11-17

-/--

INTERNATIONAL SEARCH REPORT

Interne pplication No
PCT/US 01/19032

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
T	JIANG Y ET AL: "DISCOVERY OF DIFFERENTIALLY EXPRESSED GENES IN HUMAN BREAST CANCER USING SUBTRACTED CDNA LIBRARIES AND CDNA MICROARRAYS" ONCOGENE, BASINGSTOKE, HANTS, GB, vol. 14, no. 21, 2002, pages 2270-2282, XP001075060 ISSN: 0950-9232 the whole document	1-9, 11-17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/19032

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 6,12-14,17 (in part)
because they relate to subject matter not required to be searched by this Authority, namely:
see FURTHER INFORMATION sheet PCT/ISA/210

2. [X] Claims Nos.: 10
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-9,11-17 (all in part)

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-9,11-17 (all in part)

Invention 1

An isolated polypeptide comprising a sequence defined by SEQ ID NO.56 or 98 (B511S; clone 1016F8); an expression vector comprising said polynucleotide; a host cell comprising said vector; an isolated polypeptide comprising an amino acid sequence defined by SEQ ID NO.98 or 108-116; an antibody binding said polypeptide; a diagnostic kit comprising said antibody; a fusion protein comprising said polypeptide; compositions comprising and methods using said polynucleotide or said polypeptide.

2. Claims: 1-9,11-17 (in part)

Invention 2

An isolated polypeptide comprising a sequence defined by SEQ ID NO.1; an expression vector comprising said polynucleotide; a host cell comprising said vector; an isolated polypeptide comprising an amino acid sequence encoded by SEQ ID NO.1; an antibody binding said polypeptide; a diagnostic kit comprising said antibody; a fusion protein comprising said polypeptide; compositions comprising and methods using said polynucleotide or said polypeptide.

3. Claims: 1-9,11-17 (in part)

Inventions 3-100

idem for SEQ ID NOs 2-55,57-94,102-107 respectively.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 6,12-14 and 17 comprise methods (of treatment) carried out on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.1

Claims Nos.: 6,12-14,17 (in part)

Rule 39.1(iv) PCT - Method for treatment of the human or animal body by therapy or surgery (claims 6,12-14,17)

Continuation of Box I.2

Claims Nos.: 10

Claim 10 relates to an isolated T-cell population which is defined only in terms of the method used for its preparation. As such, the subject-matter of said claim is not defined in terms which would enable the skilled person to determine what lies within the scope of said claim. In view of the wording of the claim, which render it difficult, if not impossible, to determine the matter for which protection is sought, the present application fails to comply with the clarity and conciseness requirements of Article 6 PCT (see also Rule 6.1(a) PCT) to such an extent that a meaningful search for claim 10 is impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/19032

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9933869	A	08-07-1999	US 6379951	B	30-04-2002
			US 6365348	B	02-04-2002
			AU 2010699	A	19-07-1999
			CA 2316397	A	08-07-1999
			EP 1042360	A	11-10-2000
			JP 2002507387	T	12-03-2002
			US 6410507	B	25-06-2002
			US 6432707	B	13-08-2002
			US 2001018058	A	30-08-2001
			ZA 9811800	A	23-06-1999
WO 9818945	A	07-05-1998	EP 0939824	A	08-09-1999
			JP 2001503980	T	27-03-2001
WO 9821331	A	22-05-1998	AU 7180198	A	03-06-1998
			EP 0941328	A	15-09-1999
			JP 2001527524	T	25-12-2001
			US 2002107385	A	08-08-2002
WO 9833915	A	06-08-1998	AU 6257298	A	25-08-1998
			EP 1015582	A	05-07-2000
			JP 2001509664	T	24-07-2001
WO 0175171	A	11-10-2001	AU 5307901	A	15-10-2001
			EP 1272668	A	08-01-2003
			US 2002009738	A	24-01-2002

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/98339 A2

(51) International Patent Classification[7]: C07K 14/435

(21) International Application Number: PCT/US01/19032

(22) International Filing Date: 12 June 2001 (12.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/602,877	22 June 2000 (22.06.2000)	US
09/687,507	12 October 2000 (12.10.2000)	US
09/778,381	6 February 2001 (06.02.2001)	US

(71) Applicant *(for all designated States except US)*: **CORIXA CORPORATION** [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **REED, Steven, G.** [US/US]; 2843 122nd Place N.E., Bellevue, WA 98005 (US). **XU, Jiangchun** [US/US]; 15805 S.E. 43rd Place, Bellevue, WA 98006 (US). **DILLON, Davin, C.** [US/US]; 18112 N.W. Montreux Drive, Issaquah, WA 98027 (US). **RETTER, Marc, W.** [US/US]; 33402 N.E. 43rd Place, Carnation, WA 98014 (US). **HARLOCKER, Susan, L.** [US/US]; 7522 13th Avenue W., Seattle, WA 98117 (US).

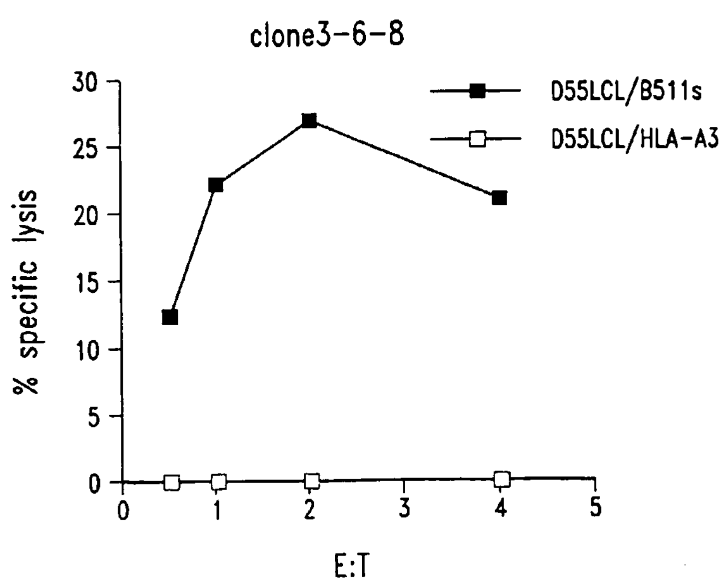
(74) Agents: **POTTER, Jane, E., R.**; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF BREAST CANCER



(57) Abstract: Compositions and methods for the therapy and diagnosis of cancer, particularly breast cancer, are disclosed. Illustrative compositions comprise one or more breast tumor polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cell that expresses such polypeptides, and T cells that are specific for cells expressing such polypeptides. The disclosed compositions are useful, for example, in the diagnosis, prevention and/or treatment of diseases, particularly breast cancer.

WO 01/98339 A2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/19032

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9933869 A	08-07-1999	US 6379951 B	30-04-2002
		US 6365348 B	02-04-2002
		AU 2010699 A	19-07-1999
		CA 2316397 A	08-07-1999
		EP 1042360 A	11-10-2000
		JP 2002507387 T	12-03-2002
		US 6410507 B	25-06-2002
		US 6432707 B	13-08-2002
		US 2001018058 A	30-08-2001
		ZA 9811800 A	23-06-1999
WO 9818945 A	07-05-1998	EP 0939824 A	08-09-1999
		JP 2001503980 T	27-03-2001
WO 9821331 A	22-05-1998	AU 7180198 A	03-06-1998
		EP 0941328 A	15-09-1999
		JP 2001527524 T	25-12-2001
		US 2002107385 A	08-08-2002
WO 9833915 A	06-08-1998	AU 6257298 A	25-08-1998
		EP 1015582 A	05-07-2000
		JP 2001509664 T	24-07-2001
WO 0175171 A	11-10-2001	AU 5307901 A	15-10-2001
		EP 1272668 A	08-01-2003
		US 2002009738 A	24-01-2002

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/98339 A2

(51) International Patent Classification[7]: C07K 14/435

(21) International Application Number: PCT/US01/19032

(22) International Filing Date: 12 June 2001 (12.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/602,877 22 June 2000 (22.06.2000) US
09/687,507 12 October 2000 (12.10.2000) US
09/778,381 6 February 2001 (06.02.2001) US

(71) Applicant *(for all designated States except US)*: **CORIXA CORPORATION** [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **REED, Steven, G.** [US/US]; 2843 122nd Place N.E., Bellevue, WA 98005 (US). **XU, Jiangchun** [US/US]; 15805 S.E. 43rd Place, Bellevue, WA 98006 (US). **DILLON, Davin, C.** [US/US]; 18112 N.W. Montreux Drive, Issaquah, WA 98027 (US). **RETTER, Marc, W.** [US/US]; 33402 N.E. 43rd Place, Carnation, WA 98014 (US). **HARLOCKER, Susan, L.** [US/US]; 7522 13th Avenue W., Seattle, WA 98117 (US).

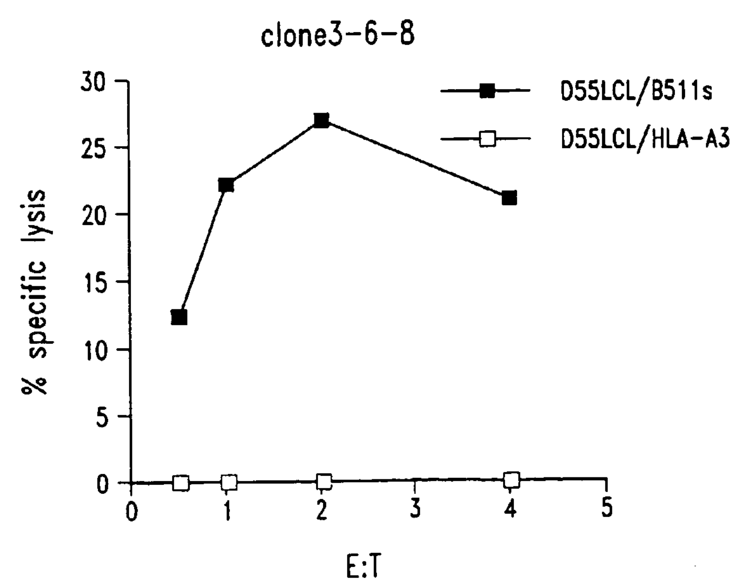
(74) Agents: **POTTER, Jane, E., R.**; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF BREAST CANCER



(57) **Abstract:** Compositions and methods for the therapy and diagnosis of cancer, particularly breast cancer, are disclosed. Illustrative compositions comprise one or more breast tumor polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cell that expresses such polypeptides, and T cells that are specific for cells expressing such polypeptides. The disclosed compositions are useful, for example, in the diagnosis, prevention and/or treatment of diseases, particularly breast cancer.

WO 01/98339 A2

IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF BREAST CANCER

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to therapy and diagnosis of cancer, such as breast cancer. The invention is more specifically related to polypeptides, comprising at least a portion of a breast tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides are useful in pharmaceutical compositions, *e.g.*, vaccines, and other compositions for the diagnosis and treatment of breast cancer.

BACKGROUND OF THE INVENTION

Breast cancer is a significant health problem for women in the United States and throughout the world. Although advances have been made in detection and treatment of the disease, breast cancer remains the second leading cause of cancer-related deaths in women, affecting more than 180,000 women in the United States each year. For women in North America, the life-time odds of getting breast cancer are now one in eight.

No vaccine or other universally successful method for the prevention or treatment of breast cancer is currently available. Management of the disease currently relies on a combination of early diagnosis (through routine breast screening procedures) and aggressive treatment, which may include one or more of a variety of treatments such as surgery, radiotherapy, chemotherapy and hormone therapy. The course of treatment for a particular breast cancer is often selected based on a variety of prognostic parameters, including an analysis of specific tumor markers. *See*, *e.g.*, Porter-Jordan and Lippman, *Breast Cancer 8*:73-100 (1994). However, the use of established markers often leads to a result that is difficult to interpret, and the high mortality observed in breast cancer patients indicates that improvements are needed in the treatment, diagnosis and prevention of the disease.

Accordingly, there is a need in the art for improved methods for therapy and diagnosis of breast cancer. The present invention fulfills these needs and further provides other related advantages.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides polynucleotide compositions comprising a sequence selected from the group consisting of:

(a) sequences provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(b) complements of the sequences provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(d) sequences that hybridize to a sequence provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118, under moderately stringent conditions;

(e) sequences having at least 75% identity to a sequence of SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(f) sequences having at least 90% identity to a sequence of SEQ ID NO: 1-97, 100, 102-107, 117 and 118; and

(g) degenerate variants of a sequence provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118.In one preferred embodiment, the polynucleotide compositions of the invention are expressed in at least about 20%, more preferably in at least about 30%, and most preferably in at least about 50% of breast tumor samples tested, at a level that is at least about 2-fold, preferably at least about 5-fold, and most preferably at least about 10-fold higher than that for normal tissues.

The present invention, in another aspect, provides polypeptide compositions comprising an amino acid sequence that is encoded by a polynucleotide sequence described above. In specific embodiments, the polypeptides of the present invention comprise at least a portion of a tumor protein that includes an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NO: 98, 99, 101, 108-116 and 119-121, and variants thereof.

In certain preferred embodiments, the polypeptides and/or polynucleotides of the present invention are immunogenic, *i.e.*, they are capable of eliciting an immune response, particularly a humoral and/or cellular immune response, as further described herein.

The present invention further provides fragments, variants and/or derivatives of the disclosed polypeptide and/or polynucleotide sequences, wherein the fragments, variants and/or derivatives preferably have a level of immunogenic activity of at least about 50%, preferably at least about 70% and more preferably at least about 90% of the level of immunogenic activity of a polypeptide sequence set forth in SEQ ID NOs: 98, 99, 101, 108-116 and 119-121 or a polypeptide sequence encoded by a polynucleotide sequence set forth in SEQ ID NOs: 1-97, 100, 102-107, 117 and 118.

The present invention further provides polynucleotides that encode a polypeptide described above, expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, the pharmaceutical compositions, *e.g.*, vaccine compositions, are provided for prophylactic or therapeutic applications. Such compositions generally comprise an immunogenic polypeptide or polynucleotide of the invention and an immunostimulant, such as an adjuvant.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a polypeptide of the present invention, or a fragment thereof; and (b) a physiologically acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Illustrative antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts and B cells.

Within related aspects, pharmaceutical compositions are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) an immunostimulant.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides

encoding such fusion proteins, typically in the form of pharmaceutical compositions, *e.g.*, vaccine compositions, comprising a physiologically acceptable carrier and/or an immunostimulant. The fusions proteins may comprise multiple immunogenic polypeptides or portions/variants thereof, as described herein, and may further comprise one or more polypeptide segments for facilitating the expression, purification and/or immunogenicity of the polypeptide(s).

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human patient, comprising administering a pharmaceutical composition described herein. The patient may be afflicted with breast cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition as recited above. The patient may be afflicted with breast cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample with T cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the

stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating $CD4^+$ and/or $CD8^+$ T cells isolated from a patient with one or more of: (i) a polypeptide comprising at least an immunogenic portion of polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer, preferably a breast cancer, in a patient comprising: (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figs. 1A and B show the specific lytic activity of a first and a second B511S-specific CTL clone, respectively, measured on autologous LCL transduced with B511S (filled squares) or HLA-A3 (open squares).

SEQ ID NO: 1 is the determined 3'cDNA sequence of 1T-5120

SEQ ID NO: 2 is the determined 3'cDNA sequence of 1T-5122

SEQ ID NO: 3 is the determined 3'cDNA sequence of 1T-5123

SEQ ID NO: 4 is the determined 3'cDNA sequence of 1T-5125

SEQ ID NO: 5 is the determined 3'cDNA sequence of 1T-5126

SEQ ID NO: 6 is the determined 3'cDNA sequence of 1T-5127

SEQ ID NO: 7 is the determined 3'cDNA sequence of 1T-5129

SEQ ID NO: 8 is the determined 3'cDNA sequence of 1T-5130

SEQ ID NO: 9 is the determined 3'cDNA sequence of 1T-5133

SEQ ID NO: 10 is the determined 3'cDNA sequence of 1T-5136

SEQ ID NO: 11 is the determined 3'cDNA sequence of 1T-5137

SEQ ID NO: 12 is the determined 3'cDNA sequence of 1T-5139

SEQ ID NO: 13 is the determined 3'cDNA sequence of 1T-5142

SEQ ID NO: 14 is the determined 3'cDNA sequence of 1T-5143

SEQ ID NO: 15 is the determined 5'cDNA sequence of 1T-5120

SEQ ID NO: 16 is the determined 5'cDNA sequence of 1T-5122

SEQ ID NO: 17 is the determined 5'cDNA sequence of 1T-5123

SEQ ID NO: 18 is the determined 5'cDNA sequence of 1T-5125

SEQ ID NO: 19 is the determined 5'cDNA sequence of 1T-5126

SEQ ID NO: 20 is the determined 5'cDNA sequence of 1T-5127

SEQ ID NO: 21 is the determined 5'cDNA sequence of 1T-5129

SEQ ID NO: 22 is the determined 5'cDNA sequence of 1T-5130

SEQ ID NO: 23 is the determined 5'cDNA sequence of 1T-5133

SEQ ID NO: 24 is the determined 5'cDNA sequence of 1T-5136

SEQ ID NO: 25 is the determined 5'cDNA sequence of 1T-5137

SEQ ID NO: 26 is the determined 5'cDNA sequence of 1T-5139

SEQ ID NO: 27 is the determined 5'cDNA sequence of 1T-5142

SEQ ID NO: 28 is the determined 5'cDNA sequence of 1T-5143

SEQ ID NO: 29 is the determined 5'cDNA sequence of 1D-4315

SEQ ID NO: 30 is the determined 5'cDNA sequence of 1D-4311

SEQ ID NO: 31 is the determined 5'cDNA sequence of 1E-4440

SEQ ID NO: 32 is the determined 5'cDNA sequence of 1E-4443

SEQ ID NO: 33 is the determined 5'cDNA sequence of 1D-4321

SEQ ID NO: 34 is the determined 5'cDNA sequence of 1D-4310

SEQ ID NO: 35 is the determined 5'cDNA sequence of 1D-4320

SEQ ID NO: 36 is the determined 5'cDNA sequence of 1E-4448

SEQ ID NO: 37 is the determined 5'cDNA sequence of 1S-5105

SEQ ID NO: 38 is the determined 5'cDNA sequence of 1S-5110

SEQ ID NO: 39 is the determined 5'cDNA sequence of 1S-5111

SEQ ID NO: 40 is the determined 5'cDNA sequence of 1S-5116

SEQ ID NO: 41 is the determined 5'cDNA sequence of 1S-5114

SEQ ID NO: 42 is the determined 5'cDNA sequence of 1S-5115

SEQ ID NO: 43 is the determined 5'cDNA sequence of 1S-5118

SEQ ID NO: 44 is the determined 5'cDNA sequence of 1T-5134

SEQ ID NO: 45 is the determined 5'cDNA sequence of 1E-4441

SEQ ID NO: 46 is the determined 5'cDNA sequence of 1E-4444

SEQ ID NO: 47 is the determined 5'cDNA sequence of 1E-4322

SEQ ID NO: 48 is the determined 5'cDNA sequence of 1S-5103

SEQ ID NO: 49 is the determined 5'cDNA sequence of 1S-5107

SEQ ID NO: 50 is the determined 5'cDNA sequence of 1S-5113

SEQ ID NO: 51 is the determined 5'cDNA sequence of 1S-5117

SEQ ID NO: 52 is the determined 5'cDNA sequence of 1S-5112

SEQ ID NO: 53 is the determined cDNA sequence of 1013E11

SEQ ID NO: 54 is the determined cDNA sequence of 1013H10

SEQ ID NO: 55 is the determined cDNA sequence of 1017C2

SEQ ID NO: 56 is the determined cDNA sequence of 1016F8

SEQ ID NO: 57 is the determined cDNA sequence of 1015F5

SEQ ID NO: 58 is the determined cDNA sequence of 1017A11

SEQ ID NO: 59 is the determined cDNA sequence of 1013A11 (also known as B537S)

SEQ ID NO: 60 is the determined cDNA sequence of 1016D8

SEQ ID NO: 61 is the determined cDNA sequence of 1016D12 (also known as B532S)

SEQ ID NO: 62 is the determined cDNA sequence of 1015E8

SEQ ID NO: 63 is the determined cDNA sequence of 1015D11 (also known as B512S)

SEQ ID NO: 64 is the determined cDNA sequence of 1012H8 (also known as B533S)

SEQ ID NO: 65 is the determined cDNA sequence of 1013C8

SEQ ID NO: 66 is the determined cDNA sequence of 1014B3

SEQ ID NO: 67 is the determined cDNA sequence of 1015B2 (also known as B536S)

SEQ ID NO: 68-71 are the determined cDNA sequences of previously identified antigens

SEQ ID NO: 72 is the determined cDNA sequence of JJ9434

SEQ ID NO: 73 is the determined cDNA sequence of B535S

SEQ ID NO: 74-88 are the determined cDNA sequences of previously identified antigens

SEQ ID NO: 89 is the determined cDNA sequence of B534S

SEQ ID NO: 90 is the determined cDNA sequence of B538S

SEQ ID NO: 91 is the determined cDNA sequence of B542S

SEQ ID NO: 92 is the determined cDNA sequence of B543S

SEQ ID NO: 93 is the determined cDNA sequence of P501S

SEQ ID NO: 94 is the determined cDNA sequence of B541S

SEQ ID NO: 95 is the full-length cDNA sequence for 1016F8 (also referred to as B511S)

SEQ ID NO: 96 is the full-length cDNA sequence for 1016D12 (also referred to as B532S)

SEQ ID NO: 97 is an extended cDNA sequence for 1012H8 (also referred to as B533S)

SEQ ID NO: 98 is the amino acid sequence for B511S

SEQ ID NO: 99 is the amino acid sequence for B532S

SEQ ID NO: 100 is the determined full-length cDNA sequence for P501S

SEQ ID NO: 101 is the amino acid sequence for P501S

SEQ ID NO: 102 is the determined cDNA sequence of clone #19605, also referred to as 1017C2, showing no significant homology to any known gene

SEQ ID NO: 103 is the determined 3' end cDNA sequence for clone #19599, showing homology to the Tumor Expression Enhanced gene

SEQ ID NO: 104 is the determined 5' end cDNA sequence for clone #19599, showing homology to the Tumor Expression Enhanced gene

SEQ ID NO: 105 is the determined cDNA sequence for clone #19607, showing homology to Stromelysin-3

SEQ ID NO: 106 is the determined cDNA sequence for clone #19601, showing homology to Collagen

SEQ ID NO: 107 is the determined cDNA sequence of clone #19606, also referred to as B546S, showing no significant homology to any known gene

SEQ ID NO: 108-116 are peptides employed in epitope mapping studies for B511S.

SEQ ID NO: 117 is the cDNA coding sequence for B543S including stop codon.

SEQ ID NO: 118 is the cDNA coding sequence for B543S without stop codon.

SEQ ID NO: 119 is the full-length amino acid sequence for B543S.

SEQ ID NO: 120 represents amino acids 1-24 of B543S.

SEQ ID NO: 121 represents amino acids 85-206 of B543S.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed generally to compositions and their use in the therapy and diagnosis of cancer, particularly breast cancer. As described further below, illustrative compositions of the present invention include, but are not restricted to, polypeptides, particularly immunogenic polypeptides, polynucleotides encoding such polypeptides, antibodies and other binding agents, antigen presenting cells (APCs) and immune system cells (*e.g.,* T cells).

The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of virology, immunology, microbiology, molecular biology and recombinant DNA techniques within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, et al. Molecular Cloning: A Laboratory Manual (2nd Edition, 1989); Maniatis et al. Molecular Cloning: A Laboratory Manual (1982); DNA Cloning: A Practical Approach, vol. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed., 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1985); Transcription and Translation (B.

Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal, A Practical Guide to Molecular Cloning (1984).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

Polypeptide Compositions

As used herein, the term "polypeptide" is used in its conventional meaning, *i.e.*, as a sequence of amino acids. The polypeptides are not limited to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide, and such terms may be used interchangeably herein unless specifically indicated otherwise. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. A polypeptide may be an entire protein, or a subsequence thereof. Particular polypeptides of interest in the context of this invention are amino acid subsequences comprising epitopes, *i.e.*, antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

Particularly illustrative polypeptides of the present invention comprise those encoded by a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-97, 100, 102-107, 117 and 118, or a sequence that hybridizes under moderately stringent conditions, or, alternatively, under highly stringent conditions, to a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-97, 100, 102-107, 117 and 118. Certain other illustrative polypeptides of the invention comprise amino acid sequences as set forth in any one of SEQ ID NOs: 98, 99, 101, 108-116 and 119-121.

The polypeptides of the present invention are sometimes herein referred to as breast tumor proteins or breast tumor polypeptides, as an indication that their identification has been based at least in part upon their increased levels of expression in

breast tumor samples. Thus, a "breast tumor polypeptide" or "breast tumor protein," refers generally to a polypeptide sequence of the present invention, or a polynucleotide sequence encoding such a polypeptide, that is expressed in a substantial proportion of breast tumor samples, for example preferably greater than about 20%, more preferably greater than about 30%, and most preferably greater than about 50% or more of breast tumor samples tested, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in normal tissues, as determined using a representative assay provided herein. A breast tumor polypeptide sequence of the invention, based upon its increased level of expression in tumor cells, has particular utility both as a diagnostic marker as well as a therapeutic target, as further described below.

In certain preferred embodiments, the polypeptides of the invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with breast cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens can be performed using methods such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ^{125}I-labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides disclosed herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of the invention that itself is immunologically reactive (*i.e.,* specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they

specifically bind to an antigen (*i.e.,* they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (*e.g., in* an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length polypeptide, *e.g.*, having greater than about 100% or 150% or more immunogenic activity.

In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- and/or C-terminal deletion (*e.g.,* 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

In another embodiment, a polypeptide composition of the invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

In another embodiment of the invention, polypeptides are provided that comprise one or more polypeptides that are capable of eliciting T cells and/or antibodies that are immunologically reactive with one or more polypeptides described herein, or one or more polypeptides encoded by contiguous nucleic acid sequences contained in the polynucleotide sequences disclosed herein, or immunogenic fragments or variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

The present invention, in another aspect, provides polypeptide fragments comprising at least about 5, 10, 15, 20, 25, 50, or 100 contiguous amino acids, or more, including all intermediate lengths, of a polypeptide compositions set forth herein, such as those set forth in SEQ ID NOs: 98, 99, 101, 108-116 and 119-121, or those encoded by a polynucleotide sequence set forth in a sequence of SEQ ID NOs: 1-97, 100, 102-107, 117 and 118.

In another aspect, the present invention provides variants of the polypeptide compositions described herein. Polypeptide variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below), along its length, to a polypeptide sequences set forth herein.

In one preferred embodiment, the polypeptide fragments and variants provide by the present invention are immunologically reactive with an antibody and/or T-cell that reacts with a full-length polypeptide specifically set for the herein.

In another preferred embodiment, the polypeptide fragments and variants provided by the present invention exhibit a level of immunogenic activity of at least about 50%, preferably at least about 70%, and most preferably at least about 90% or more of that exhibited by a full-length polypeptide sequence specifically set forth herein.

A polypeptide "variant," as the term is used herein, is a polypeptide that typically differs from a polypeptide specifically disclosed herein in one or more substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the above polypeptide sequences of the invention and evaluating their immunogenic activity as described herein and/or using any of a number of techniques well known in the art.

For example, certain illustrative variants of the polypeptides of the invention include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other illustrative variants

include variants in which a small portion (*e.g.,* 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

In many instances, a variant will contain conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. As described above, modifications may be made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, immunogenic variant or portion of a polypeptide of the invention, one skilled in the art will typically change one or more of the codons of the encoding DNA sequence according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

TABLE 1

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are:

isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (–0.4); threonine (–0.7); serine (–0.8); tryptophan (–0.9); tyrosine (–1.3); proline (–1.6); histidine (–3.2); glutamate (–3.5); glutamine (–3.5); aspartate (–3.5); asparagine (–3.5); lysine (–3.9); and arginine (–4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (–0.4); proline (–0.5 ± 1); alanine (–0.5); histidine (–0.5); cysteine (–1.0); methionine (–1.3); valine (–1.5); leucine (–1.8); isoleucine (–1.8); tyrosine (–2.3); phenylalanine (–2.5); tryptophan (–3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl-methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydropathic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.,* poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be "identical" if the sequence of amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two

sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenes pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS 5*:151-153; Myers, E.W. and Muller W. (1988) *CABIOS 4*:11-17; Robinson, E.D. (1971) *Comb. Theor 11*:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol. 4*:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA 80*:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402

and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide may be a fusion polypeptide that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to

desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the polypeptide.

Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is expressed as a recombinant polypeptide, allowing the production of increased levels, relative to a non-fused polypeptide, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene 40*:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA 83*:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements

responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide as described herein together with an unrelated immunogenic protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see*, for example, Stoute et al. *New Engl. J. Med.*, *336*:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is derived from a Mycobacterium sp., such as a Mycobacterium tuberculosis-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is described in U.S. Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid. MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid sequence of MTB32A have been described (for example, U.S. Patent Application 60/158,585; see also, Skeiky *et al.*, *Infection and Immun.* (1999) 67:3998-4007, incorporated herein by reference). C-terminal fragments of the MTB32A coding sequence express at high levels and remain as a soluble polypeptides throughout the purification process. Moreover, Ra12 may enhance the immunogenicity of heterologous immunogenic polypeptides with which it is fused. One preferred Ra12 fusion polypeptide comprises a 14 KD C-terminal fragment corresponding to amino acid residues 192 to 323 of MTB32A. Other preferred Ra12 polynucleotides generally comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide. Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a

sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium Haemophilus influenza B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.,* the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemaglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene 43*:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology 10*:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated

into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

Yet another illustrative embodiment involves fusion polypeptides, and the polynucleotides encoding them, wherein the fusion partner comprises a targeting signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention, when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced in vivo stimulation of $CD4^+$ T-cells specific for the polypeptide.

Polypeptides of the invention are prepared using any of a variety of well known synthetic and/or recombinant techniques, the latter of which are further described below. Polypeptides, portions and other variants generally less than about 150 amino acids can be generated by synthetic means, using techniques well known to those of ordinary skill in the art. In one illustrative example, such polypeptides are synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc. 85*:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

In general, polypeptide compositions (including fusion polypeptides) of the invention are isolated. An "isolated" polypeptide is one that is removed from its original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are also purified, *e.g.*, are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

Polynucleotide Compositions

The present invention, in other aspects, provides polynucleotide compositions. The terms "DNA" and "polynucleotide" are used essentially

interchangeably herein to refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA molecule does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be understood by those skilled in the art, the polynucleotide compositions of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

As will be also recognized by the skilled artisan, polynucleotides of the invention may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules may include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.,* an endogenous sequence that encodes a polypeptide/protein of the invention or a portion thereof) or may comprise a sequence that encodes a variant or derivative, preferably and immunogenic variant or derivative, of such a sequence.

Therefore, according to another aspect of the present invention, polynucleotide compositions are provided that comprise some or all of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-97, 100, 102-107, 117 and 118, complements of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-97, 100, 102-107, 117 and 118, and degenerate variants of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-97, 100, 102-107, 117 and 118. In certain preferred

embodiments, the polynucleotide sequences set forth herein encode immunogenic polypeptides, as described above.

In other related embodiments, the present invention provides polynucleotide variants having substantial identity to the sequences disclosed herein in SEQ ID NOs: 1-97, 100, 102-107, 117 and 118, for example those comprising at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide sequence of this invention using the methods described herein, (*e.g.*, BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

Typically, polynucleotide variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the immunogenicity of the polypeptide encoded by the variant polynucleotide is not substantially diminished relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompasses homologous genes of xenogenic origin.

In additional embodiments, the present invention provides polynucleotide fragments comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 10, 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

In another embodiment of the invention, polynucleotide compositions are provided that are capable of hybridizing under moderate to high stringency conditions to

a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-60°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated, such as by altering the salt content of the hybridization solution and/or the temperature at which the hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those described above, with the exception that the temperature of hybridization is increased, *e.g.*, to 60-65°C or 65-70°C.

In certain preferred embodiments, the polynucleotides described above, *e.g.*, polynucleotide variants, fragments and hybridizing sequences, encode polypeptides that are immunologically cross-reactive with a polypeptide sequence specifically set forth herein. In other preferred embodiments, such polynucleotides encode polypeptides that have a level of immunogenic activity of at least about 50%, preferably at least about 70%, and more preferably at least about 90% of that for a polypeptide sequence specifically set forth herein.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

When comparing polynucleotide sequences, two sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenes pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS 5*:151-153; Myers, E.W. and Muller W. (1988) *CABIOS 4*:11-17; Robinson, E.D. (1971) *Comb. Theor 11*:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol. 4*:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA 80*:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.,* gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.,* the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Therefore, in another embodiment of the invention, a mutagenesis approach, such as site-specific mutagenesis, is employed for the preparation of immunogenic variants and/or derivatives of the polypeptides described herein. By this approach, specific modifications in a polypeptide sequence can be made through mutagenesis of the underlying polynucleotides that encode them. These techniques provides a straightforward approach to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the polynucleotide.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the properties of the polynucleotide itself, and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the immunogenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the

art, and are widely used to create variants of both polypeptides and polynucleotides. For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis provides a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the

teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kuby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

In another approach for the production of polypeptide variants of the present invention, recursive sequence recombination, as described in U.S. Patent No. 5,837,458, may be employed. In this approach, iterative cycles of recombination and screening or selection are performed to "evolve" individual polynucleotide variants of the invention having, for example, enhanced immunogenic activity.

In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth herein, or to any continuous portion of the sequences, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various

factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus,

hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

According to another embodiment of the present invention, polynucleotide compositions comprising antisense oligonucleotides are provided. Antisense oligonucleotides have been demonstrated to be effective and targeted inhibitors of protein synthesis, and, consequently, provide a therapeutic approach by which a disease can be treated by inhibiting the synthesis of proteins that contribute to the disease. The efficacy of antisense oligonucleotides for inhibiting protein synthesis is well established. For example, the synthesis of polygalactauronase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1, striatal $GABA_A$ receptor and human EGF (Jaskulski *et al.*, Science. 1988 Jun 10;240(4858):1544-6; Vasanthakumar and Ahmed, Cancer Commun. 1989;1(4):225-32; Peris *et al.*, Brain Res Mol Brain Res. 1998 Jun 15;57(2):310-20; U. S. Patent 5,801,154; U.S. Patent 5,789,573; U. S. Patent 5,718,709 and U.S. Patent 5,610,288). Antisense constructs have also been described that inhibit and can be used to treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S. Patent 5,591,317 and U. S. Patent 5,783,683).

Therefore, in certain embodiments, the present invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof. In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein.

Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence and determination of secondary structure, T_m, binding energy, and relative stability. Antisense compositions may be selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell. Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which are substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations can be performed, for example, using v.4 of the OLIGO primer analysis software and/or the BLASTN 2.0.5 algorithm software (Altschul *et al.,* Nucleic Acids Res. 1997, 25(17):3389-402).

The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, Nucleic Acids Res. 1997 Jul 15;25(14):2730-6). It has been demonstrated that several molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane.

According to another embodiment of the invention, the polynucleotide compositions described herein are used in the design and preparation of ribozyme molecules for inhibiting expression of the tumor polypeptides and proteins of the present invention in tumor cells. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, Proc Natl Acad Sci U S A. 1987 Dec;84(24):8788-92; Forster and Symons, Cell. 1987 Apr 24;49(2):211-20). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, Cell. 1981 Dec;27(3 Pt 2):487-96; Michel and Westhof, J Mol Biol. 1990 Dec 5;216(3):585-610; Reinhold-Hurek and Shub, Nature.

1992 May 14;357(6374):173-6). This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action (Woolf *et al.*, Proc Natl Acad Sci U S A. 1992 Aug 15;89(16):7305-9). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis δ virus, group I intron or RNaseP RNA (in association with an RNA

guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* Nucleic Acids Res. 1992 Sep 11;20(17):4559-65. Examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz, Biochemistry 1989 Jun 13;28(12):4929-33; Hampel *et al.,* Nucleic Acids Res. 1990 Jan 25;18(2):299-304 and U. S. Patent 5,631,359. An example of the hepatitis δ virus motif is described by Perrotta and Been, Biochemistry. 1992 Dec 1;31(47):11843-52; an example of the RNaseP motif is described by Guerrier-Takada *et al.,* Cell. 1983 Dec;35(3 Pt 2):849-57; Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, Cell. 1990 May 18;61(4):685-96; Saville and Collins, Proc Natl Acad Sci U S A. 1991 Oct 1;88(19):8826-30; Collins and Olive, Biochemistry. 1993 Mar 23;32(11):2795-9); and an example of the Group I intron is described in (U. S. Patent 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see *e.g.*, Int. Pat. Appl. Publ. No. WO 92/07065; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which

enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. No. WO 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells Ribozymes expressed from such promoters have been shown to function in mammalian cells. Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

In another embodiment of the invention, peptide nucleic acids (PNAs) compositions are provided. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, Antisense Nucleic Acid Drug Dev. 1997 7(4) 431-37). PNA is able to be utilized in a number methods that traditionally have used RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (*Trends Biotechnol* 1997 Jun;15(6):224-9). As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen *et al.*, *Science* 1991 Dec 6;254(5037):1497-500; Hanvey *et al.*, Science. 1992 Nov 27;258(5087):1481-5; Hyrup and Nielsen, Bioorg Med Chem. 1996 Jan;4(1):5-23). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc or Fmoc protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used.

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton *et al.*, Bioorg Med Chem. 1995 Apr;3(4):437-45). The manual protocol lends itself to the production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines

can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography, providing yields and purity of product similar to those observed during the synthesis of peptides.

Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications of PNAs (for example, Norton *et al.*, Bioorg Med Chem. 1995 Apr;3(4):437-45; Petersen *et al.*, J Pept Sci. 1995 May-Jun;1(3):175-83; Orum *et al.*, Biotechniques. 1995 Sep;19(3):472-80; Footer *et al.*, Biochemistry. 1996 Aug 20;35(33):10673-9; Griffith *et al.*, Nucleic Acids Res. 1995 Aug 11;23(15):3003-8; Pardridge *et al.*, Proc Natl Acad Sci U S A. 1995 Jun 6;92(12):5592-6; Boffa *et al.*, Proc Natl Acad Sci U S A. 1995 Mar 14;92(6):1901-5; Gambacorti-Passerini *et al.*, Blood. 1996 Aug 15;88(4):1411-7; Armitage *et al.*, Proc Natl Acad Sci U S A. 1997 Nov 11;94(23):12320-5; Seeger *et al.*, Biotechniques. 1997 Sep;23(3):512-7). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (Anal Chem. 1993 Dec 15;65(24):3545-9) and Jensen *et al.* (Biochemistry. 1997 Apr 22;36(16):5072-7). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* using BIAcore™ technology.

Other applications of PNAs that have been described and will be apparent to the skilled artisan include use in DNA strand invasion, antisense inhibition, mutational analysis, enhancers of transcription, nucleic acid purification, isolation of transcriptionally active genes, blocking of transcription factor binding, genome cleavage, biosensors, *in situ* hybridization, and the like.

Polynucleotide Identification, Characterization and Expression

Polynucleotides compositions of the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.,* expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA 93*:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA 94*:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR™, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will

dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR™ amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Any of a number of other template dependent processes, many of which are variations of the PCR ™ amplification technique, are readily known and available in the art. Illustratively, some such methods include the ligase chain reaction (referred to as LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.*, a tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.,* by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence. One such amplification technique is inverse PCR (*see* Triglia et al., *Nucl. Acids Res. 16*:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic. 1*:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids.*

Res. 19:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.,* NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science 269*:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and

translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York. N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, any of a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be

used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of .beta.-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem. 264*:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J. 6*:307-311. Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J. 3*:1671-1680; Broglie, R. et al. (1984) *Science 224*:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ. 17*:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, Autographa californica nuclear polyhedrosis virus

(AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. et al. (1994) *Proc. Natl. Acad. Sci. 91* :3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci. 81*:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ. 20*:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation. glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, COS, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell 11*:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1990) *Cell 22*:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. 77*:3567-70); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol. 150*:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in

place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci. 85*:8047-51). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol. 55*:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include, for example, membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med. 158*:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen. San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues

facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* *3*:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* *12*:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* *85*:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

<u>Antibody Compositions, Fragments Thereof and Other Binding Agents</u>

According to another aspect, the present invention further provides binding agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a tumor polypeptide disclosed herein, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," “immunogically bind,” and/or is “immunologically reactive” to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant (K_d) of the interaction, wherein a smaller K_d represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates

depend on the concentrations of the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" (K_{on}) and the "off rate constant" (K_{off}) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of K_{off}/K_{on} enables cancellation of all parameters not related to affinity, and is thus equal to the dissociation constant K_d. See, generally, Davies et al. (1990) Annual Rev. Biochem. 59:439-473.

An "antigen-binding site," or "binding portion" of an antibody refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs."

Binding agents may be further capable of differentiating between patients with and without a cancer, such as breast cancer, using the representative assays provided herein. For example, antibodies or other binding agents that bind to a tumor protein will preferably generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.*, blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein

for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with and without the disease will be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol. 6*:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as

described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) each comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including the "$F(ab')_2$ " fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using recombinant techniques known in the art. The Fv fragment includes a non-covalent V_H::V_L heterodimer including an antigen-binding site which retains much of the antigen recognition and binding capabilities of the native antibody molecule.

Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem 15:2706-2710; and Ehrlich et al. (1980) Biochem 19:4091-4096.

A single chain Fv ("sFv") polypeptide is a covalently linked V_H::V_L heterodimer which is expressed from a gene fusion including V_H- and V_L-encoding genes linked by a peptide-encoding linker. Huston et al. (1988) Proc. Nat. Acad. Sci. USA 85(16):5879-5883. A number of methods have been described to discern chemical structures for converting the naturally aggregated--but chemically separated--light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Pat. No. 4,946,778, to Ladner et al.

Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRS and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (*e.g.*, a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRS. Within FRs, certain amino residues and certain structural features are very highly conserved. In this regard, all V region sequences contain an

internal disulfide loop of around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of FRs which influence the folded shape of the CDR loops into certain "canonical" structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al. (1991) Nature 349:293-299; Lobuglio et al. (1989) Proc. Nat. Acad. Sci. USA 86:4220-4224; Shaw et al. (1987) J Immunol. 138:4534-4538; and Brown et al. (1987) Cancer Res. 47:3577-3583), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann et al. (1988) Nature 332:323-327; Verhoeyen et al. (1988) Science 239:1534-1536; and Jones et al. (1986) Nature 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, *e.g.*, a rodent heavy or light chain V region, with human FR residues in order to provide a xenogeneic molecule comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the antigen-binding surface. Davies et al. (1990) Ann. Rev. Biochem. 59:439-473. Thus, antigen binding specificity can be preserved in a humanized antibody only wherein the CDR structures, their interaction with each other, and their interaction with the rest of

the V region domains are carefully maintained. By using veneering techniques, exterior (*e.g.*, solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S. Government Printing Office, 1987), updates to the Kabat database, and other accessible U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody molecule of interest are compared with corresponding FR sequences of human variable domains obtained from the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are thus designed to retain the murine CDR residues, the residues substantially adjacent to the CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues believed to participate in non-covalent (*e.g.*, electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs which are believed to influence the "canonical" tertiary structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences which combine the CDRs of both the heavy and light chain of a murine antigen-binding site into human-appearing FRs that can be used to transfect

mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y, ^{123}I, ^{125}I, ^{131}I, ^{186}Re, ^{188}Re, ^{211}At, and ^{212}Bi. Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diptheria toxin, cholera toxin, gelonin, Pseudomonas exotoxin, Shigella toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

T Cell Compositions

The present invention, in another aspect, provides T cells specific for a tumor polypeptide disclosed herein, or for a variant or derivative thereof. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a polypeptide, polynucleotide encoding a polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide of interest. Preferably, a tumor polypeptide or polynucleotide of the invention is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a polypeptide of the present invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res. 54*:1065-1070, 1994. Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a tumor polypeptide (100 ng/ml - 100 μg/ml, preferably 200 ng/ml - 25 μg/ml) for 3 - 7 days will typically result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as

measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.*, TNF or IFN-γ) is indicative of T cell activation (*see* Coligan et al., Current Protocols in Immunology, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a tumor polypeptide, polynucleotide or polypeptide-expressing APC may be $CD4^+$ and/or $CD8^+$. Tumor polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, $CD4^+$ or $CD8^+$ T cells that proliferate in response to a tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of the tumor polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

Pharmaceutical Compositions

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed herein in pharmaceutically-acceptable carriers for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will be understood that, if desired, a composition as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from

host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Therefore, in another aspect of the present invention, pharmaceutical compositions are provided comprising one or more of the polynucleotide, polypeptide, antibody, and/or T-cell compositions described herein in combination with a physiologically acceptable carrier. In certain preferred embodiments, the pharmaceutical compositions of the invention comprise immunogenic polynucleotide and/or polypeptide compositions of the invention for use in prophylactic and theraputic vaccine applications. Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Generally, such compositions will comprise one or more polynucleotide and/or polypeptide compositions of the present invention in combination with one or more immunostimulants.

It will be apparent that any of the pharmaceutical compositions described herein can contain pharmaceutically acceptable salts of the polynucleotides and polypeptides of the invention. Such salts can be prepared, for example, from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.*, salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (*e.g.*, sodium, potassium, lithium, ammonium, calcium and magnesium salts).

In another embodiment, illustrative immunogenic compositions, *e.g.*, vaccine compositions, of the present invention comprise DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems 15*:143-198, 1998, and references cited therein. Appropriate polynucleotide expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal). Alternatively, bacterial delivery systems may involve

the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the present invention can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to a subject. A number of illustrative retroviral systems have been described (*e.g.*, U.S. Pat. No. 5,219,740; Miller and Rosman (1989) BioTechniques 7:980-990; Miller, A. D. (1990) Human Gene Therapy 1:5-14; Scarpa et al. (1991) Virology 180:849-852; Burns et al. (1993) Proc. Natl. Acad. Sci. USA 90:8033-8037; and Boris-Lawrie and Temin (1993) Cur. Opin. Genet. Develop. 3:102-109.

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham (1986) J. Virol. 57:267-274; Bett et al. (1993) J. Virol. 67:5911-5921; Mittereder et al. (1994) Human Gene Therapy 5:717-729; Seth et al. (1994) J. Virol. 68:933-940; Barr et al. (1994) Gene Therapy 1:51-58; Berkner, K. L. (1988) BioTechniques 6:616-629; and Rich et al. (1993) Human Gene Therapy 4:461-476).

Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, *e.g.*, U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al. (1988) Molec. Cell. Biol. 8:3988-3996; Vincent et al. (1990) Vaccines 90 (Cold Spring Harbor Laboratory Press); Carter, B. J. (1992) Current Opinion in Biotechnology 3:533-539; Muzyczka, N. (1992) Current Topics in Microbiol. and Immunol. 158:97-129; Kotin, R. M. (1994) Human Gene Therapy 5:793-801; Shelling and Smith (1994) Gene Therapy 1:165-169; and Zhou et al. (1994) J. Exp. Med. 179:1867-1875.

Additional viral vectors useful for delivering the polynucleotides encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

A vaccinia-based infection/transfection system can be conveniently used to provide for inducible, transient expression or coexpression of one or more polypeptides described herein in host cells of an organism. In this particular system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide or polynucleotides of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, *e.g.*, Elroy-Stein and Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst et al. Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an Avipox vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and

therefore are not infective in mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, *e.g.*, WO 91/12882; WO 89/03429; and WO 92/03545.

Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

Moreover, molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al. J. Biol. Chem. (1993) 268:6866-6869 and Wagner et al. Proc. Natl. Acad. Sci. USA (1992) 89:6099-6103, can also be used for gene delivery under the invention.

Additional illustrative information on these and other known viral-based delivery systems can be found, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA 86*:317-321, 1989; Flexner et al., *Ann. N.Y. Acad. Sci. 569*:86-103, 1989; Flexner et al., *Vaccine 8*:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques 6*:616-627, 1988; Rosenfeld et al., *Science 252*:431-434, 1991; Kolls et al., *Proc. Natl. Acad. Sci. USA 91*:215-219, 1994; Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA 90*:11498-11502, 1993; Guzman et al., *Circulation 88*:2838-2848, 1993; and Guzman et al., *Cir. Res. 73*:1202-1207, 1993.

In certain embodiments, a polynucleotide may be integrated into the genome of a target cell. This integration may be in the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the polynucleotide may be stably maintained in the cell as a separate, episomal segment of DNA. Such polynucleotide segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. The manner in which the expression construct is delivered to a cell and

where in the cell the polynucleotide remains is dependent on the type of expression construct employed.

In another embodiment of the invention, a polynucleotide is administered/delivered as "naked" DNA, for example as described in Ulmer et al., *Science 259*:1745-1749, 1993 and reviewed by Cohen, *Science 259*:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In still another embodiment, a composition of the present invention can be delivered via a particle bombardment approach, many of which have been described. In one illustrative example, gas-driven particle acceleration can be achieved with devices such as those manufactured by Powderject Pharmaceuticals PLC (Oxford, UK) and Powderject Vaccines Inc. (Madison, WI), some examples of which are described in U.S. Patent Nos. 5,846,796; 6,010,478; 5,865,796; 5,584,807; and EP Patent No. 0500 799. This approach offers a needle-free delivery approach wherein a dry powder formulation of microscopic particles, such as polynucleotide or polypeptide particles, are accelerated to high speed within a helium gas jet generated by a hand held device, propelling the particles into a target tissue of interest.

In a related embodiment, other devices and methods that may be useful for gas-driven needle-less injection of compositions of the present invention include those provided by Bioject, Inc. (Portland, OR), some examples of which are described in U.S. Patent Nos. 4,790,824; 5,064,413; 5,312,335; 5,383,851; 5,399,163; 5,520,639 and 5,993,412.

According to another embodiment, the pharmaceutical compositions described herein will comprise one or more immunostimulants in addition to the immunogenic polynucleotide, polypeptide, antibody, T-cell and/or APC compositions of this invention. An immunostimulant refers to essentially any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. One preferred type of immunostimulant comprises an adjuvant. Many adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bortadella pertussis* or *Mycobacterium tuberculosis* derived proteins.

Certain adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN-γ, TNFα, IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A, together with an aluminum salt. MPL® adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science 273*:352, 1996. Another preferred adjuvant comprises a saponin,

such as Quil A, or derivatives thereof, including QS21 and QS7. (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins . Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β-escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamelar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^R to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL® adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D-MPL® adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 is disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.,* SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Enhanzyn®) (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula

(I): $HO(CH_2CH_2O)_n$-A-R,

wherein, n is 1-50, A is a bond or –C(O)-, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein n is between 1 and 50, preferably 4-24, most preferably 9; the R component is C_{1-50}, preferably C_4-C_{20} alkyl and most preferably C_{12} alkyl, and A is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-steoryl ether, polyoxyethylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12^{th} edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

According to another embodiment of this invention, an immunogenic composition described herein is delivered to a host via antigen presenting cells (APCs),

such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.,* matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature 392*:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman and Levy, *Ann. Rev. Med. 50*:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med. 4*:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNFα to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNFα,

CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcγ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.,* CD54 and CD11) and costimulatory molecules (*e.g.,* CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.,* vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.,* a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier

will typically vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, mucosal, intravenous, intracranial, intraperitoneal, subcutaneous and intramuscular administration.

Carriers for use within such pharmaceutical compositions are biocompatible, and may also be biodegradable. In certain embodiments, the formulation preferably provides a relatively constant level of active component release. In other embodiments, however, a more rapid rate of release immediately upon administration may be desired. The formulation of such compositions is well within the level of ordinary skill in the art using known techniques. Illustrative carriers useful in this regard include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other illustrative delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

In another illustrative embodiment, biodegradable microspheres (*e.g.*, polylactate polyglycolate) are employed as carriers for the compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344, 5,407,609 and 5,942,252. Modified hepatitis B core protein carrier systems. such as described in WO/99 40934, and references cited therein, will also be useful for many applications. Another illustrative carrier/delivery system employs a carrier comprising particulate-protein complexes, such as those described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

The pharmaceutical compositions of the invention will often further comprise one or more buffers (*e.g.*, neutral buffered saline or phosphate buffered

saline), carbohydrates (*e.g.*, glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate.

The pharmaceutical compositions described herein may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are typically sealed in such a way to preserve the sterility and stability of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

The development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation, is well known in the art, some of which are briefly discussed below for general purposes of illustration.

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (see, for example, Mathiowitz *et al.*, Nature 1997 Mar 27;386(6623):410-4; Hwang *et al.*, Crit Rev Ther Drug Carrier Syst 1998;15(3):243-84; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451). Tablets, troches, pills, capsules and the like may also contain any of a variety of additional components, for example, a binder, such as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent,

such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations will contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared is such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even

intraperitoneally. Such approaches are well known to the skilled artisan, some of which are further described, for example, in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S. Patent 5,399,363. In certain embodiments, solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations generally will contain a preservative to prevent the growth of microorganisms.

Illustrative pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (for example, see U. S. Patent 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and/or by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

In one embodiment, for parenteral administration in an aqueous solution, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will

be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. Moreover, for human administration, preparations will of course preferably meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

In another embodiment of the invention, the compositions disclosed herein may be formulated in a neutral or salt form. Illustrative pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective.

The carriers can further comprise any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the

lungs *via* nasal aerosol sprays has been described, *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212. Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, J Controlled Release 1998 Mar 2;52(1-2):81-7) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871) are also well-known in the pharmaceutical arts. Likewise, illustrative transmucosal drug delivery in the form of a polytetrafluoroetheylene support matrix is described in U. S. Patent 5,780,045.

In certain embodiments, liposomes, nanocapsules, microparticles, lipid particles, vesicles, and the like, are used for the introduction of the compositions of the present invention into suitable host cells/organisms. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like. Alternatively, compositions of the present invention can be bound, either covalently or non-covalently, to the surface of such carrier vehicles.

The formation and use of liposome and liposome-like preparations as potential drug carriers is generally known to those of skill in the art (see for example, Lasic, Trends Biotechnol 1998 Jul;16(7):307-21; Takakura, Nippon Rinsho 1998 Mar;56(3):691-5; Chandran *et al.*, Indian J Exp Biol. 1997 Aug;35(8):801-9; Margalit, Crit Rev Ther Drug Carrier Syst. 1995;12(2-3):233-61; U.S. Patent 5,567,434; U.S. Patent 5,552,157; U.S. Patent 5,565,213; U.S. Patent 5,738,868 and U.S. Patent 5,795,587, each specifically incorporated herein by reference in its entirety).

Liposomes have been used successfully with a number of cell types that are normally difficult to transfect by other procedures, including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, J Biol Chem. 1990 Sep 25;265(27):16337-42; Muller *et al.*, DNA Cell Biol. 1990 Apr;9(3):221-9). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, various drugs, radiotherapeutic agents, enzymes, viruses, transcription factors, allosteric effectors and the like, into a variety of cultured cell lines and animals. Furthermore, he use of liposomes does not appear to be associated with autoimmune responses or unacceptable toxicity after systemic delivery.

In certain embodiments, liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs).

Alternatively, in other embodiments, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (see, for example, Quintanar-Guerrero *et al.*, Drug Dev Ind Pharm. 1998 Dec;24(12):1113-28). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) may be designed using polymers able to be degraded *in vivo*. Such particles can be made as described, for example, by Couvreur *et al.,* Crit Rev Ther Drug Carrier Syst. 1988;5(1):1-20; zur Muhlen *et al.*, Eur J Pharm Biopharm. 1998 Mar;45(2):149-55; Zambaux *et al.* J Controlled Release. 1998 Jan 2;50(1-3):31-40; and U. S. Patent 5,145,684.

Cancer Therapeutic Methods

In further aspects of the present invention, the pharmaceutical compositions described herein may be used for the treatment of cancer, particularly for the immunotherapy of breast cancer. Within such methods, the pharmaceutical compositions described herein are administered to a patient, typically a warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Accordingly, the above pharmaceutical compositions may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. As discussed above, administration of the pharmaceutical compositions may be by any suitable method, including administration by intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune

response-modifying agents (such as polypeptides and polynucleotides as provided herein).

Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as $CD8^+$ cytotoxic T lymphocytes and $CD4^+$ T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with immunoreactive polypeptides or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies

have shown that cultured effector cells can be induced to grow in vivo and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (*see*, for example, Cheever et al., *Immunological Reviews 157*:177, 1997).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitary, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 25 μg to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical

outcome (*e.g.*, more frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

Cancer Detection and Diagnostic Compositions, Methods and Kits

In general, a cancer may be detected in a patient based on the presence of one or more breast tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies) obtained from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer such as breast cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample. Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also indicative of the presence or absence of a cancer. In general, a breast tumor sequence should be present at a level that is at least three fold higher in tumor tissue than in normal tissue

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory, 1988. In general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a

binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G, protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length breast tumor proteins and polypeptide portions thereof to which the binding agent binds, as described above.

The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681. The binding agent may be immobilized on the solid support using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10 µg, and preferably about 100 ng to about 1 µg, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding partner (*see*, *e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13)..

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with breast cancer. Preferably, the contact time is sufficient to achieve a level of binding that is at least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of

time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, such as breast cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity)

that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1 μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a tumor protein in a biological sample. Within certain methods, a biological sample comprising $CD4^+$ and/or $CD8^+$ T cells isolated from a patient is incubated with a tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (*e.g.*, 5 - 25 µg/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of tumor polypeptide to serve as a control. For $CD4^+$ T cells, activation is preferably detected by evaluating proliferation of the T cells. For $CD8^+$ T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a tumor protein in a biological sample. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes that specifically hybridize to a

polynucleotide encoding a tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a tumor protein of the invention that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence as disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, *51*:263, 1987; Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another embodiment, the compositions described herein may be used as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of

reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a

polynucleotide encoding a tumor protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

ISOLATION AND CHARACTERIZATION OF BREAST TUMOR POLYPEPTIDES

This Example describes the isolation of breast tumor polypeptides from a breast tumor cDNA library.

A human breast tumor cDNA expression library was constructed from a pool of breast tumor poly A^+ RNA from three patients using a Superscript Plasmid System for cDNA Synthesis and Plasmid Cloning kit (BRL Life Technologies, Gaithersburg, MD 20897) following the manufacturer's protocol. Specifically, breast tumor tissues were homogenized with polytron (Kinematica, Switzerland) and total RNA was extracted using Trizol reagent (BRL Life Technologies) as directed by the manufacturer. The poly A^+ RNA was then purified using a Qiagen oligotex spin column mRNA purification kit (Qiagen, Santa Clarita, CA 91355) according to the manufacturer's protocol. First-strand cDNA was synthesized using the NotI/Oligo-dT18 primer. Double-stranded cDNA was synthesized, ligated with EcoRI/BstX I adaptors (Invitrogen, Carlsbad, CA) and digested with NotI. Following size fractionation with Chroma Spin-1000 columns (Clontech, Palo Alto, CA 94303), the cDNA was ligated into the EcoRI/NotI site of pCDNA3.1 (Invitrogen, Carlsbad, CA) and transformed into ElectroMax *E. coli* DH10B cells (BRL Life Technologies) by electroporation.

Using the same procedure, a normal human breast cDNA expression library was prepared from a pool of four normal breast tissue specimens. The cDNA libraries were characterized by determining the number of independent colonies, the percentage of clones that carried insert, the average insert size and by sequence analysis.

The breast tumor library contained 1.14 x 10^7 independent colonies, with more than 90% of clones having a visible insert and the average insert size being 936 base pairs. The normal breast cDNA library contained 6 x 10^6 independent colonies, with 83% of clones having inserts and the average insert size being 1015 base pairs. Sequencing analysis showed both libraries to contain good complex cDNA clones that were synthesized from mRNA, with minimal rRNA and mitochondrial DNA contamination sequencing.

cDNA library subtraction was performed using the above breast tumor and normal breast cDNA libraries, as described by Hara *et al.* (*Blood*, *84*:189-199, 1994) with some modifications. Specifically, a breast tumor-specific subtracted cDNA library was generated as follows. Normal breast cDNA library (70 µg) was digested with EcoRI, NotI, and SfuI, followed by a filling-in reaction with DNA polymerase Klenow fragment. After phenol-chloroform extraction and ethanol precipitation, the DNA was dissolved in 100 µl of H_2O, heat-denatured and mixed with 100 µl (100 µg) of Photoprobe biotin (Vector Laboratories, Burlingame, CA), the resulting mixture was irradiated with a 270 W sunlamp on ice for 20 minutes. Additional Photoprobe biotin (50 µl) was added and the biotinylation reaction was repeated. After extraction with butanol five times, the DNA was ethanol-precipitated and dissolved in 23 µl H_2O to form the driver DNA.

To form the tracer DNA, 10 µg breast tumor cDNA library was digested with BamHI and XhoI, phenol chloroform extracted and passed through Chroma spin-400 columns (Clontech). Following ethanol precipitation, the tracer DNA was dissolved in 5 µl H_2O. Tracer DNA was mixed with 15 µl driver DNA and 20 µl of 2 x hybridization buffer (1.5 M NaCl/10 mM EDTA/50 mM HEPES pH 7.5/0.2% sodium dodecyl sulfate), overlaid with mineral oil, and heat-denatured completely. The sample was immediately transferred into a 68 ^{0}C water bath and incubated for 20 hours (long hybridization [LH]). The reaction mixture was then subjected to a streptavidin treatment followed by phenol/chloroform extraction. This process was repeated three more times. Subtracted DNA was precipitated, dissolved in 12 µl H_2O, mixed with 8 µl driver DNA and 20 µl of 2 x hybridization buffer, and subjected to a hybridization at 68 ^{0}C for 2 hours (short hybridization [SH]). After removal of biotinylated double-

stranded DNA, subtracted cDNA was ligated into BamHI/XhoI site of chloramphenicol resistant pBCSK$^+$ (Stratagene, La Jolla, CA 92037) and transformed into ElectroMax *E. coli* DH10B cells by electroporation to generate a breast tumor specific subtracted cDNA library.

To analyze the subtracted cDNA library, plasmid DNA was prepared from 100 independent clones, randomly picked from the subtracted breast tumor specific library and characterized by DNA sequencing with a Perkin Elmer/Applied Biosystems Division Automated Sequencer Model 373A (Foster City, CA). Thirty-eight distinct cDNA clones were found in the subtracted breast tumor-specific cDNA library. The determined 3' cDNA sequences for 14 of these clones are provided in SEQ ID NO: 1-14, with the corresponding 5' cDNA sequences being provided in SEQ ID NO: 15-28, respectively. The determined one strand (5' or 3') cDNA sequences for the remaining clones are provided in SEQ ID NO: 29-52. Comparison of these cDNA sequences with known sequences in the gene bank using the EMBL and GenBank databases (Release 97) revealed no significant homologies to the sequences provided in SEQ ID NO: 3, 10, 17, 24 and 45-52. The sequences provided in SEQ ID NO: 1, 2, 4-9, 11-16, 18-23, 25-41, 43 and 44 were found to show at least some degree of homology to known human genes. The sequence of SEQ ID NO: 42 was found to show some homology to a known yeast gene.

cDNA clones isolated in the breast subtraction described above were colony PCR amplified and their mRNA expression levels in breast tumor, normal breast and various other normal tissues were determined using microarray technology (Synteni, Fremont, CA). Briefly, the PCR amplification products were dotted onto slides in an array format, with each product occupying a unique location in the array. mRNA was extracted from the tissue sample to be tested, reverse transcribed, and fluorescent-labeled cDNA probes were generated. The microarrays were probed with the labeled cDNA probes, the slides scanned and fluorescence intensity was measured. This intensity correlates with the hybridization intensity.

Data was analyzed using GEMTOOLS Software. Twenty one distinct cDNA clones were found to be over-expressed in breast tumor and expressed at low levels in all normal tissues tested. The determined partial cDNA sequences for these

clones are provided in SEQ ID NO: 53-73. Comparison of the sequences of SEQ ID NO: 53, 54 and 68-71 with those in the gene bank as described above, revealed some homology to previously identified human genes. No significant homologies were found to the sequences of SEQ ID NO: 55-67, 72 (referred to as JJ 9434) and 73 (referred to as B535S). In further studies, full length cDNA sequences were obtained for the clones 1016F8 (SEQ ID NO: 56; also referred to as B511S) and 1016D12 (SEQ ID NO: 61; also referred to as B532S), and an extended cDNA sequence was obtained for 1012H8 (SEQ ID NO: 64; also referred to as B533S). These cDNA sequences are provided in SEQ ID NO: 95-97, respectively, with the corresponding amino acid sequences for B511S and B532S being provided in SEQ ID NO: 98 and 99, respectively.

Analysis of the expression of B511S in breast tumor tissues and in a variety of normal tissues (skin, PBMC, intestine, breast, stomach, liver, kidney, fetal tissue, adrenal gland, salivary gland, spinal cord, large intestine, small intestine, bone marrow, brain, heart, colon and pancreas) by microarray, northern analysis and real time PCR, demonstrated that B511S is over-expressed in breast tumors, and normal breast, skin and salivary gland, with expression being low or undetectable in all other tissues tested.

Analysis of the expression of B532S in breast tumor tissue and in a variety of normal tissues (breast, PBMC, esophagus, HMEC, spinal cord, bone, thymus, brain, bladder, colon, liver, lung, skin, small intestine, stomach, skeletal muscle, pancreas, aorta, heart, spleen, kidney, salivary gland, bone marrow and adrenal gland) by microarray, Northern analysis and real time PCR, demonstrated that B532S is over-expressed in 20-30% of breast tumors with expression being low or undetectable in all other tissues tested.

In a further experiment, cDNA fragments were obtained from two subtraction libraries derived by conventional subtraction, as described above and analyzed by DNA microarray. In one instance the tester was derived from primary breast tumors, referred to as Breast Subtraction 2, or BS2. In the second instance, a metastatic breast tumor was employed as the tester, referred to as Breast Subtraction 3, or BS3. Drivers consisted of normal breast.

cDNA fragments from these two libraries were submitted as templates for DNA microarray analysis, as described above. DNA chips were analyzed by hybridizing with fluorescent probes derived from mRNA from both tumor and normal tissues. Analysis of the data was accomplished by creating three groups from the sets of probes, referred to as breast tumor/mets, normal non-breast tissues, and metastatic breast tumors. Two comparisons were performed using the modified Gemtools analysis. The first comparison was to identify templates with elevated expression in breast tumors. The second was to identify templates not recovered in the first comparison that yielded elevated expression in metastatic breast tumors. An arbitrary level of increased expression (mean of tumor expression versus the mean of normal tissue expression) was set at approximately 2.2.

In the first round of comparison to identify over-expression in breast tumors, two novel gene sequences were identified, hereinafter referred to as B534S and B538S (SEQ ID NO: 89 and 90, respectively), together with six sequences that showed some degree of homology to previously identified genes (SEQ ID NO: 74-79). The sequences of SEQ ID NO: 75 and 76 were subsequently determined to be portions of B535S (SEQ ID NO: 73). In a second comparison to identify elevated expression in metastatic breast tumors, five novel sequences were identified, hereinafter referred to as B535S, B542S, B543S, P501S and B541S (SEQ ID NO: 73 and 91-94, respectively), as well as nine gene sequences that showed some homology to known genes (SEQ ID NO: 80-88). Clones B534S and B538S (SEQ ID NO: 89 and 90) were shown to be over-expressed in both breast tumors and metastatic breast tumors.

The cDNA sequence of B543S (SEQ ID NO: 92) was found to contain a 206 amino acid open reading frame (ORF) encoded by nucleotides 71-691. The cDNA sequence of the B543S coding sequence with stop codon is provided in SEQ ID NO: 117, with the cDNA sequence of the B543S coding sequence without stop codon being provided in SEQ ID NO: 118. The corresponding full-length amino acid sequence is provided in SEQ ID NO: 119. This amino acid sequence was analyzed using the computer algorithm PSORT II in order to identify putative transmembrane domains. A single transmembrane domain was identified located at residues 8-24. SEQ ID NO: 120 and 121 represent amino acids 1-24 and 85-206, respectively, of B543S.

In a subsequent series of studies, 457 clones from Breast Subtraction 2 were analyzed by microarray on Breast Chip 3. As described above, a first comparison to identify over-expression in breast tumors over normal non-breast tissues was performed. This analysis yielded six cDNA clones that demonstrated elevated expression in breast tumor over normal non-breast tissues. Two of these clones, referred to as 1017C2 (SEQ ID NO: 102) and B546S (SEQ ID NO: 107) do not share significant homology to any known genes. Clone B511S also showed over-expression in breast tumor, which was previously described as 1016F8, with the determined cDNA sequence provided in SEQ ID NO: 95 and the amino acid sequence provided in SEQ ID NO: 98. The remaining four clones over-expressed in breast tumor were found to share some degree of homology to Tumor Expression Enhanced Gene (SEQ ID NO: 103 and 104) Stromelysin-3 (SEQ ID NO: 105) or Collagen (SEQ ID NO: 106).

In the second comparison to determine genes with elevated expression in metastatic breast tumors over non-breast normal tissues, a profile similar to the first comparison was derived. The two putatively novel clones, 1017C2 and B546S, SEQ ID NO: 102 and 107, respectively, were overexpressed in metastatic breast tumors. In addition, Tumor Expression Enhanced Gene and B511S also showed elevated expression in metastatic breast tumors.

As described in U.S. Patent Application No. 08/806,099, filed February 25, 1997, the antigen P501S was isolated by subtracting a prostate tumor cDNA library with a normal pancreas cDNA library and with three genes found to be abundant in a previously subtracted prostate tumor specific cDNA library: human glandular kallikrein, prostate specific antigen (PSA), and mitochondria cytochrome C oxidase subunit II. The determined full-length cDNA sequence for P501S is provided in SEQ ID NO: 100, with the corresponding amino acid sequence being provided in SEQ ID NO: 101. Expression of P501S in breast tumor was examined by microarray analysis. Over-expression was found in prostate tumor, breast tumor and metastatic breast tumor, with negligible to low expression being seen in normal tissues. This data suggests that P501S may be over-expressed in various breast tumors as well as in prostate tumors.

EXAMPLE 2

GENERATION OF HUMAN CD8+ CYTOTOXIC T-CELLS THAT RECOGNIZE ANTIGEN PRESENTING CELLS EXPRESSING BREAST TUMOR ANTIGENS

This Example illustrates the generation of T cells that recognize target cells expressing the antigen B511S, also known as 1016-F8 (SEQ ID NO: 95). Human CD8+ T cells were primed *in-vitro* to the B511S gene product using dendritic cells infected with a recombinant vaccinia virus engineered to express B511S as follows (also see Yee et al., Journal of Immunology (1996) 157 (9):4079-86). Dendritic cells (DC) were generated from peripheral blood derived monocytes by differentiation for 5 days in the presence of 50 μg/ml GMCSF and 30 μg/ml IL-4. DC were harvested, plated in wells of a 24-well plate at a density of 2 x 10^5 cells/well and infected for 12 hours with B511S expressing vaccinia at a multiplicity of infection of 5. DC were then matured overnight by the addition of 3 μg/ml CD40-Ligand and UV irradiated at 100μW for 10 minutes. CD8+ T cells were isolated using magnetic beads, and priming cultures were initiated in individual wells (typically in 24 wells of a 24-well plate) using 7 x 10^5 CD8+ T cells and 1 x 10^6 irradiated CD8-depleted PBMC. IL-7 at 10 ng/ml was added to cultures at day 1. Cultures were re-stimulated every 7-10 days using autologous primary fibroblasts retrovirally transduced with B511S and the costimulatory molecule B7.1. Cultures were supplemented at day 1 with 15 I.U. of IL-2. Following 4 such stimulation cycles, CD8+ cultures were tested for their ability to specifically recognize autologous fibroblasts transduced with B511S using an interferon-γ Elispot assay (see Lalvani et al J. Experimental Medicine (1997) 186:859-965). Briefly, T cells from individual microcultures were added to 96-well Elispot plates that contained autologous fibroblasts transduced to express either B511S or as a negative control antigen EGFP, and incubated overnight at 37° C; wells also contained IL-12 at 10 ng/ml. Cultures were identified that specifically produced interferon-γ only in response to B511S transduced fibroblasts; such lines were further expanded and also cloned by limiting dilution on autologous B-LCL retrovirally transduced with B511S. Lines and clones were identified that could specifically recognize autologous B-LCL transduced with B511S but not autologous B-LCL transduced with the control antigens EGFP or HLA-A3. An example demonstrating the ability of human CTL cell lines

derived from such experiments to specifically recognize and lyse B511S expressing targets is presented in Figure 1.

EXAMPLE 3

PREPARATION AND CHARACTERIZATION OF ANTIBODIES AGAINST BREAST TUMOR POLYPEPTIDES

Polyclonal antibodies against the breast tumor antigens B511S and B532S were prepared as follows.

The breast tumor antigen expressed in an *E. coli* recombinant expression system was grown overnight in LB broth with the appropriate antibiotics at 37 °C in a shaking incubator. The next morning, 10 ml of the overnight culture was added to 500 ml to 2x YT plus appropriate antibiotics in a 2L-baffled Erlenmeyer flask. When the Optical Density (at 560 nm) of the culture reached 0.4-0.6, the cells were induced with IPTG (1 mM). Four hours after induction with IPTG, the cells were harvested by centrifugation. The cells were then washed with phosphate buffered saline and centrifuged again. The supernatant was discarded and the cells were either frozen for future use or immediately processed. Twenty ml of lysis buffer was added to the cell pellets and vortexed. To break open the *E. coli* cells, this mixture was then run through the French Press at a pressure of 16,000 psi. The cells were then centrifuged again and the supernatant and pellet were checked by SDS-PAGE for the partitioning of the recombinant protein. For proteins that localized to the cell pellet, the pellet was resuspended in 10 mM Tris pH 8.0, 1% CHAPS and the inclusion body pellet was washed and centrifuged again. This procedure was repeated twice more. The washed inclusion body pellet was solubilized with either 8 M urea or 6 M guanidine HCl containing 10 mM Tris pH 8.0 plus 10 mM imidazole. The solubilized protein was added to 5 ml of nickel-chelate resin (Qiagen) and incubated for 45 min to 1 hour at room temperature with continuous agitation. After incubation, the resin and protein mixture were poured through a disposable column and the flow through was collected. The column was then washed with 10-20 column volumes of the solubilization buffer. The antigen was then eluted from the column using 8M urea, 10 mM Tris pH 8.0 and

300 mM imidazole and collected in 3 ml fractions. A SDS-PAGE gel was run to determine which fractions to pool for further purification.

As a final purification step, a strong anion exchange resin such as HiPrepQ (Biorad) was equilibrated with the appropriate buffer and the pooled fractions from above were loaded onto the column. Antigen was eluted off the column with a increasing salt gradient. Fractions were collected as the column was run and another SDS-PAGE gel was run to determine which fractions from the column to pool. The pooled fractions were dialyzed against 10 mM Tris pH 8.0. The protein was then vialed after filtration through a 0.22 micron filter and the antigens were frozen until needed for immunization.

Four hundred micrograms of breast tumor antigen was combined with 100 micrograms of muramyldipeptide (MDP). Every four weeks rabbits were boosted with 100 micrograms mixed with an equal volume of Incomplete Freund's Adjuvant (IFA). Seven days following each boost, the animal was bled. Sera was generated by incubating the blood at 4 ^{o}C for 12-24 hours followed by centrifugation.

Ninety-six well plates were coated with breast tumor antigen by incubating with 50 microliters (typically 1 microgram) of recombinant protein at 4 ^{o}C for 20 hours. 250 microliters of BSA blocking buffer was added to the wells and incubated at room temperature for 2 hours. Plates were washed 6 times with PBS/0.01% Tween. Rabbit sera was diluted in PBS. Fifty microliters of diluted sera was added to each well and incubated at room temperature for 30 min. Plates were washed as described above before 50 microliters of goat anti-rabbit horse radish peroxidase (HRP) at a 1:10000 dilution was added and incubated at room temperature for 30 min. Plates were again washed as described above and 100 microliters of TMB microwell peroxidase substrate was added to each well. Following a 15 min incubation in the dark at room temperature, the colorimetric reaction was stopped with 100 microliters of 1N H_2SO_4 and read immediately at 450 nm. The polyclonal antibodies prepared against B511S and B532S showed immunoreactivity to B511S and B532S, respectively.

Immunohistochemical (IHC) analysis of B511S expression in breast cancer and normal breast specimens was performed as follows. Paraffin-embedded

formal fixed tissue was sliced into 8 micron sections. Steam heat induced epitope retrieval (SHIER) in 0.1 M sodium citrate buffer (pH 6.0) was used for optimal staining conditions. Sections were incubated with 10% serum/PBS for 5 minutes. Primary antibody was added to each section for 25 min at indicated concentrations followed by a 25 min incubation with either an anti-rabbit or anti-mouse biotinylated antibody. Endogenous peroxidase activity was blocked by three 1.5 min incubations with hydrogen peroxide. The avidin biotin complex/horseradish peroxidase (ABC/HRP) system was used along with DAB chromagen to visualize antigen expression. Slides were counterstained with hematoxylin.

A summary of real-time PCR and immunohistochemical analysis of B511S expression in normal and breast tumor tissues is presented in Table 2 below. B511S expression was detected in normal breast and breast tumor tissues, as well as in skin. B511S protein expression was also detected in colon, but neither protein nor mRNA was detected in a panel of normal tissues that includes kidney, brain, liver, lung, heart and bone marrow.

TABLE 2

Tissue type	IHC staining	mRNA analysis
Breast tumor	Positive	Positive
Normal breast	Positive	Positive
Skin	Positive (apocrine only)	Negative
Colon	Positive	Negative
Kidney	Negative	Negative
Brain	Negative	Negative
Liver	Negative	Negative
Lung	Negative	Negative
Heart	Negative	Negative
Bone marrow	Negative	Negative

EXAMPLE 4

EPITOPE MAPPING OF THE BREAST TUMOR ANTIGEN B511S

Rabbit polyclonal anti-sera raised against *E. coli* derived full-length B511S recombinant protein (in the form of a thiol reduction fusion protein, referred to as B511S-Trx) and against truncated B511S as described above, together with human monoclonal antibodies against B511S, were tested for epitope recognition against a series of overlapping 15-mer peptides that correspond to the full-length B511S amino acid sequence (SEQ ID NO: 98). The truncated form of B511S, referred to as B511S-A, consisted of amino acids 21-90 of SEQ ID NO: 98 plus a 6x histidine tag. The sequences of the 15-mer peptides, corresponding to amino acids 1-15, 11-25, 21-35, 31-45, 41-55, 51-65, 61-75, 76-90 and 71-85 of B511S, are provided in SEQ ID NO: 108-116, respectively.

To prepare the human monoclonal antibodies, transgenic mice that contain human immunoglobulin gene loci for the production of human monoclonal antibodies (Abgenix Inc., Fremont, CA) were immunized with *E. coli* derived B511S-A protein and subsequently used for splenic B cell fusions to generate hybridomas. For polyclonal antibody purification, rabbit anti-B511S-A sera (referred to as 739/142) was passed over a B511S-sepharose affinity column. The rabbit anti-B511S-Trx sera 542/27 was passed over a Trx affinity column, whereas the anti-B511S-Trx sera 542/28 was passed over a Trx column followed by a B511S affinity column. All antibodies were eluted with a salt buffer containing 0.5M NaCl and 20mM phosphate, followed by

an acid elution step using 0.2M glycine, pH 2.3. Purified antibodies were neutralized by the addition of 1M Tris, pH 8 and buffer exchanged into PBS.

For ELISA analysis, 96 well plates were coated by adding either B511S peptides or recombinant B511S proteins (all antigens diluted to 2 μg/ml), and incubating for 60 minutes at 37 °C. After coating, plates were blocked with 1% BSA in PBS for 2 hours at room temperature followed by incubation overnight at 4 °C. Plates were washed five times with PBS + 0.1% Tween 20 followed by the addition of either polyclonal sera at 1 μg/ml or hybridoma supernatants undiluted or diluted at 1:5, and incubation for 30 minutes at room temperature. Plates were washed as above and HRP-linked secondary antibodies (donkey anti-rabbit Ig-HRP for the polyclonal sera and mouse anti-human Ig-HRP for the hybridoma supernatants) were added and incubated for 30 minutes at room temperature, followed by a final washing as above. TMB peroxidase substrate was added and incubated 15 minutes at room temperature in the dark. The reaction was stopped by the addition of 1N H_2SO_4 and the OD was read at 450 nM.

The purified polyclonal anti-B511S sera was found to recognize peptides spanning amino acids 21 to 35 (SEQ ID NO: 110); amino acids 61-75 (SEQ ID NO: 114), amino acids 71 to 85 (SEQ ID NO: 116), and amino acids 76 to 90 (SEQ ID NO: 115) of the full-length B511S protein. The human hybridoma 1.6 secreted monoclonal antibody that recognized amino acids 76-90 of B511S (SEQ ID NO: 115), while both the 1.17 and 1.26 clones secreted monoclonal antibodies that recognized amino acids 71-85 and 76-90. Hybridoma 1.21 secreted monoclonal antibody that weakly bound amino acids 71-85 but clearly bound the B511S-A recombinant protein.

FACS analysis revealed that anti-B511S-Trx sera recognizes B511S/HEK stable transfectants, where anti-B511S-A sera does not recognize the same cells, suggesting that recognition of peptide 21-35 (SEQ ID NO: 110) is required for the detection of B511S surface expression.

EXAMPLE 5

PROTEIN EXPRESSION OF BREAST TUMOR ANTIGENS

This example describes the expression and purification of the breast tumor antigen B511S in mammalian cells.

Full-length B511S (SEQ ID NO: 95) was subcloned into the mammalian expression vectors pCEP4 (Invitrogen). This construct was transfected into HEK293 cells (ATCC) using Fugene 6 reagent (Roche). Briefly, the HEK cells were plated at a density of 100,000 cells/ml in DMEM (Gibco) containing 10% FBS (Hyclone) and grown overnight. The following day, 2 ul of Fugene 6 was added to 100 ul of DMEM containing no FBS and incubated for 15 minutes at room temperature. The Fugene 6/DMEM mixture was added to 1 ug of B511S/pCEP4 plasmid DNA and incubated for 15 minutes at room temperature. The Fugene/DNA mix was then added to the HEK293 cells and incubated for 48-72 hours at 37 °C with 7% CO_2. Cells were rinsed with PBS, then collected and pelleted by centrifugation.

For Western blot analysis, whole cell lysates were generated by incubating the cells in Triton-X100 containing lysis buffer for 30 minutes on ice. Lysates were then cleared by centrifugation at 10,000 rpm for 5 minutes at 4 °C. Samples were diluted with SDS_PAGE loading buffer containing beta-mercaptoethanol, and boiled for 10 minutes prior to loading the SDS_PAGE gel. Proteins were transferred to nitrocellulose and probed using Protein A purified anti-B511S rabbit polyclonal sera (prepared as described above) at a concentration of 1 ug/ml. The blot was revealed with a goat anti-rabbit Ig coupled to HRP followed by incubation in ECL substrate. Expression of B511S was detected in the the HEK293 lysates transfected with B511S, but not in control HEK293 cells transfected with vector alone.

For FACS analysis, cells were washed further with ice cold staining buffer. Next the cells were incubated for 30 minutes on ice with 10 ug/ml of Protein A purified anti-B511S polyclonal sera. The cells were washed 3 times with staining buffer and then incubated with a 1:100 dilution of a goat anti-rabbit Ig (H+L)-FITC reagent (Southern Biotechnology) for 30 minutes on ice. Following 3 washes, the cells were

resuspended in staining buffer containing Propidium Iodide (PI), a vital stain that allows for identification of permeable cells, and then analyzed by FACS. Surface expression of B511S was observed.

EXAMPLE 6

SYNTHESIS OF POLYPEPTIDES

Polypeptides may be synthesized on an Perkin Elmer/Applied Biosystems Division 430A peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation, binding to an immobilized surface, or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray or other types of mass spectrometry and by amino acid analysis.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

What is Claimed:

1. An isolated polynucleotide comprising a sequence selected from the group consisting of:

(a) sequences provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(b) complements of the sequences provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(d) sequences that hybridize to a sequence provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118, under moderately stringent conditions;

(e) sequences having at least 75% identity to a sequence of SEQ ID NO: 1-97, 100, 102-107, 117 and 118;

(f) sequences having at least 90% identity to a sequence of SEQ ID NO: 1-97, 100, 102-107, 117 and 118; and

(g) degenerate variants of a sequence provided in SEQ ID NO: 1-97, 100, 102-107, 117 and 118.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) sequences provided in SEQ ID NO: 98, 99, 101, 108-116 and 119-121;

(b) sequences encoded by a polynucleotide of claim 1;

(c) sequences having at least 70% identity to a sequence encoded by a polynucleotide of claim 1; and

(d) sequences having at least 90% identity to a sequence encoded by a polynucleotide of claim 1.

3. An expression vector comprising a polynucleotide of claim 1 operably linked to an expression control sequence.

4. A host cell transformed or transfected with an expression vector according to claim 3.

5. An isolated antibody, or antigen-binding fragment thereof, that specifically binds to a polypeptide of claim 2.

6. A method for detecting the presence of a cancer in a patient, comprising the steps of:

(a) obtaining a biological sample from the patient;

(b) contacting the biological sample with a binding agent that binds to a polypeptide of claim 2;

(c) detecting in the sample an amount of polypeptide that binds to the binding agent; and

(d) comparing the amount of polypeptide to a predetermined cut-off value and therefrom determining the presence of a cancer in the patient.

7. A fusion protein comprising at least one polypeptide according to claim 2.

8. An oligonucleotide that hybridizes to a sequence recited in SEQ ID NO: 1-97, 100, 102-107, 117 and 118 under moderately stringent conditions.

9. A method for stimulating and/or expanding T cells specific for a tumor protein, comprising contacting T cells with at least one component selected from the group consisting of:

(a) polypeptides according to claim 2;

(b) polynucleotides according to claim 1; and

(c) antigen-presenting cells that express a polypeptide according to claim 2,

under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells.

10. An isolated T cell population, comprising T cells prepared according to the method of claim 9.

11. A composition comprising a first component selected from the group consisting of physiologically acceptable carriers and immunostimulants, and a second component selected from the group consisting of:

(a) polypeptides according to claim 2;

(b) polynucleotides according to claim 1;

(c) antibodies according to claim 5;

(d) fusion proteins according to claim 7;

(e) T cell populations according to claim 10; and

(f) antigen presenting cells that express a polypeptide according to claim 2.

12. A method for stimulating an immune response in a patient, comprising administering to the patient a composition of claim 11.

13. A method for the treatment of a cancer in a patient, comprising administering to the patient a composition of claim 11.

14. A method for determining the presence of a cancer in a patient, comprising the steps of:

(a) obtaining a biological sample from the patient;

(b) contacting the biological sample with an oligonucleotide according to claim 8;

(c) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; and

(d) compare the amount of polynucleotide that hybridizes to the oligonucleotide to a predetermined cut-off value, and therefrom determining the presence of the cancer in the patient.

15. A diagnostic kit comprising at least one oligonucleotide according to claim 8.

16. A diagnostic kit comprising at least one antibody according to claim 5 and a detection reagent, wherein the detection reagent comprises a reporter group.

17. A method for inhibiting the development of a cancer in a patient, comprising the steps of:

(a) incubating CD4+ and/or CD8+ T cells isolated from a patient with at least one component selected from the group consisting of: (i) polypeptides according to claim 2; (ii) polynucleotides according to claim 1; and (iii) antigen presenting cells that express a polypeptide of claim 2, such that T cell proliferate;

(b) administering to the patient an effective amount of the proliferated T cells,

thereby inhibiting the development of a cancer in the patient.

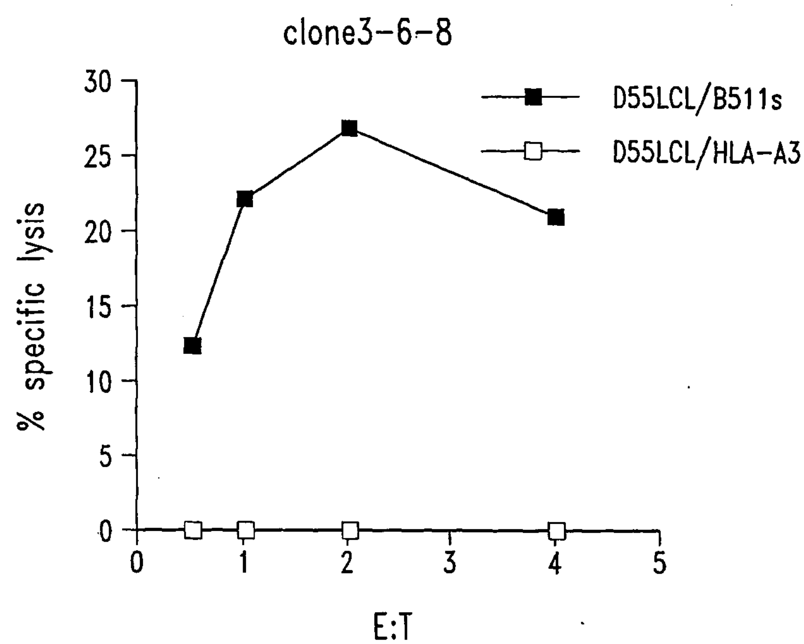

clone3-6-8
% specific lysis
30
25
20
15
10
5
0
0
1
2
3
4
5
E:T
D55LCL/B511s
D55LCL/HLA-A3

Fig. 1A

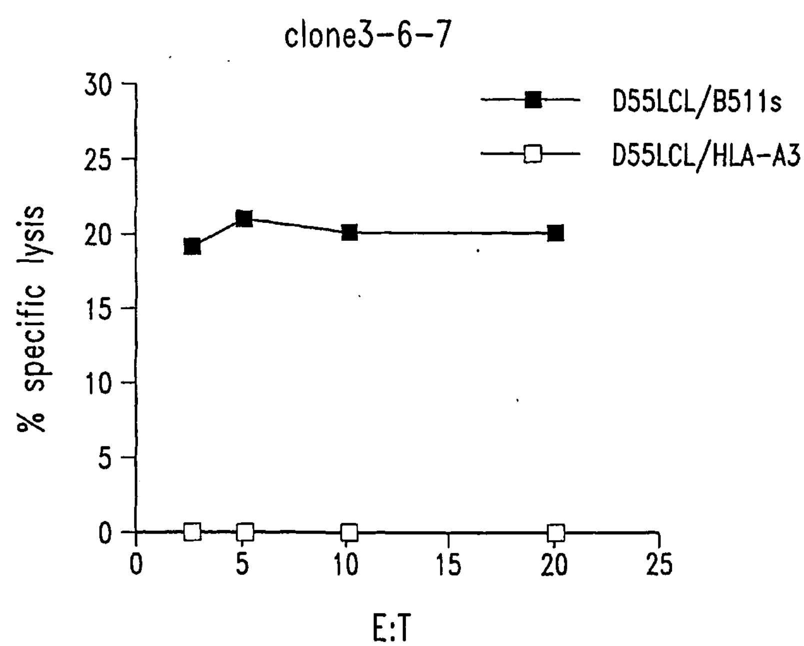

clone3-6-7
% specific lysis
30
25
20
15
10
5
0
0
5
10
15
20
25
E:T
D55LCL/B511s
D55LCL/HLA-A3

Fig. 1B

SEQUENCE LISTING

<110> Corixa Corporation
Reed, Steven G.
Xu, Jiangchun
Dillon, Davin C.
Retter, Marc W.
Harlocker, Susan L.

<120> COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF BREAST CANCER

<130> 210121.44603PC

<140> PCT
<141> 2001-06-12

<160> 121

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 402
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(402)
<223> n = A,T,C or G

<400> 1
```
tttttttttt tttttaggag aactgaatca aacagatttt attcaacttt ttagatgagg      60
aaaacaaatn atacgaaatn ngtcataaga aatgctttct tataccacta tctcaaacca     120
ctttcaatat tttacaaaat gctcacgcag caaatatgaa aagctncaac acttcccttt     180
gttaacttgc tgcaatnaat gcaactttaa canacataca aatttcttct gtatcttaaa     240
agttnaatta ctaattttaa tgatnttnct caagatnttt attcatatac ttttaatgac     300
tcnttgccna tacatacnta ttttctttac ttttttttta cnatnggcca acagctttca     360
ngcagnccnc aaaaatctta ccggttaatt acacggggtt gt                        402
```

<210> 2
<211> 424
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(424)
<223> n = A,T,C or G

<400> 2
```
tttttttttt ttttttaaag gtacacattt ctttttcatt ctgtttnatg cagcaaataa      60
ttcgttggca tcttctctgt gatgggcagc ttgctaaaat tanactcagg ccccttagct     120
ncatttccaa ctnagcccac gctttcaacc nngccnaaca aagaaaatca gttngggtta     180
aattctttgc tgganacaaa gaactacatt cctttgtaaa tnatgctttg tttgctctgt     240
gcaaacncag attgaaggga anaagganac ttntggggac ggaaacaact ngnagaagca     300
gganccgccc agggncattt cctcaccatg cttaatcttg cnctcacttg cngggcacca     360
ttaaacttgg tgcaaaaggc gcaattggtg nanggaaccc cacaccttcc ttaaaaagca     420
gggc                                                                  424
```

<210> 3
<211> 421
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(421)
<223> n = A,T,C or G

<400> 3
```
tttttttttt tttttcccaa tttaaaaaag cctttttcat acttcaatta caccanactt      60
aatnatttca tgagtaaatc ngacattatt atttnaaaat ttgcatattt aaaatttgna     120
tcanttactt ccagactgtt tgcanaatga agggaggatc actcaagngc tgatctcnca     180
ctntctgcag tctnctgtcc tgtgcccggn ctaatggatc gacactanat ggacagntcn     240
cagatcttcc gttcttntcc cttccccaat ttcncaccnc tccccttctt ncccggatcn     300
tttggggaca tgntaatttt gcnatcctta aaccctgccc gccangggtc ccnanctcag     360
gggtggttaa tgttcgncng gcttnttgac cncctgcgcc ctttnantcc naaccccaag     420
c                                                                     421
```

<210> 4
<211> 423
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(423)
<223> n = A,T,C or G

<400> 4
```
ttttttttatt ttttttcta tttntnntat ttnntgnggt tcctgtgtgt aattagnang      60
tgtgtatgcg tangtacnta tgtntgcata tttaacctgt tncctttcca tttttaaaat     120
aaaatctcaa natngtantt ggttnatggg agtaaanaga gactatngat naattttaac     180
atggacacng tgaaatgtag ccgctnatca ntttaaaact tcattttgaa ggcctttttnc    240
cctccnaata aaaatnccng gccctactgg gttaagcaac attgcatntc taaagaaacc     300
acatgcanac nagttaaacc tgtgnactgg tcangcaaac cnanntggaa nanaagggnn     360
ttcnccccan ggacantcng aattttttta acaaattacn atnccccccc ngggggagcc     420
tgt                                                                   423
```

<210> 5
<211> 355
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(355)
<223> n = A,T,C or G

<400> 5
```
acgaccacct natttcgtat ctttcaactc ttttcgaccg gacctcttat tcggaagcgt      60
tccaggaaga caggtctcaa cttagggatc agatcacgtt atcaacgctc tgggatcgct     120
gcaacctggc acttcaagga agtgcaccga tnacgtctag accggccaac acagatctag     180
aggtggccaa ctgatcactg taggagctga ctggcaanan tcaaccgggc cccaaccnag     240
agtgaccaan acnaccattn aggatcaccc acaggcactc ctcgtcctag ggccaaccna     300
ccaaacggct ggccaatggg ggggtttaat atttggttna aaaattgatt ttaaa          355
```

<210> 6
<211> 423
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(423)
<223> n = A,T,C or G

<400> 6
```
ttttttttt tttttggaca ggaagtaaaa tttattggtn antattaana gggggcagc      60
acattggaag ccctcatgan tgcagggccc gccacttgtc cagagggcca cnattgggga    120
tgtacttaac cccacagccn tctgggatna gccgcttttc agccaccatn tcttcaaatt    180
catcagcatt aaacttggta aanccccact tctttaagat ntgnatcttc tggcggccag    240
naaacttgaa cttggccctg cgcagggcct caatcacatg ctccttgttc tgcagcttgg    300
tgcgnaagga cntaatnact tggccnatgt gaaccctggc cacantgccc tggggctttc    360
caaaggcacc tcgcaagcct ntttggancc tgnccgcccc ngcacaggga caacatcttg    420
ttt                                                                  423
```

<210> 7
<211> 410
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(410)
<223> n = A,T,C or G

<400> 7
```
ttcgcactgg ctaaaacaaa ccgccttgca aagttngaaa aatttatcaa tggaccaaat     60
aatgctcata tccnacaagt tggtgaccgt tnttatnata aaaaaatgta tnatgctcct    120
nanttgttgt acaataatgt tccaatttng gacnttcggc atctaccctg gttcacctgg    180
gtaaatatca ggcagctttt gatggggcta ggaaagctaa cagtactcga acatgggaaa    240
gaggtctgct tcgccngtgt anatgggaaa naattccgtc ttgctcngat ttgtggactt    300
catattgttg tacatgcaga tgaatnngaa gaacttgtca actactatca ggatcgtggc    360
tttttnnaaa agctnatcac catgttggaa gcggcactng gacttgagcg               410
```

<210> 8
<211> 274
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(274)
<223> n = A,T,C or G

<400> 8
```
ttttttttt tttttaggtc atacatattt tttattataa canatatntg tatatacata      60
taatatatgt gtatatatcc acgtgtgtgt gtgtgtatca aaaacaacan aantttagtg    120
atctatatct ntngctcaca tatgcatggg agataccagt aaaaaataag tnaatctcca    180
taatatgttt taaaactcan anaaatcnga gagactnaaa gaaaacgttn atcannatga    240
ttgtngataa tcttgaanaa tnacnaaaac atat                                274
```

<210> 9
<211> 322
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(322)
<223> n = A,T,C or G

<400> 9
tttttttttt ttttgtgcct tattgcaccg gcnanaactt ctagcactat attaaactca 60
ataagagtga taagtgtgaa aatccttgcc ttctctttaa tcttaatgna naggcatctg 120
gtttttcacc attaantgta ataatggctn tatgtatttt tatnnatggt cttnatggag 180
ttaaaaaagt tttcctctnt ccctngttat ctaanagttt tnatcaaaaa tgggtataat 240
atttngttca gtacttttnc ctgcacctat agatatgatn ctgttatttt ttcttcttng 300
cctnnanata tgatggatna ca 322

<210> 10
<211> 425
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(425)
<223> n = A,T,C or G

<400> 10
tttttttttt tttttattct gcagccatta aatgctgaac actagatnct tatttgtgga 60
ggtcacaaaa taagtacaga atatnacaca cgccctgccc ataaaaagca cagctcccag 120
ttctatattt acaatatctc tggaattcca ccttcccttc taatttgact aatatttctg 180
cttctcaggc agcagcgcct tctggcaacc ataagaacca acntgnggac taggtcggtg 240
ggccaaggat caggaaacag aanaatggaa gnagccccn tgacnctatt aanctntnaa 300
actatctnaa ctgctagttt tcaggcttta aatcatgtaa natacgtgtc cttnttgctg 360
caaccggaag catcctagat ggtacactct ctccaggtgc caggaaaaga tcccaaatng 420
caggn 425

<210> 11
<211> 424
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(424)
<223> n = A,T,C or G

<400> 11
ttttnttant tttttttancc nctnntccnn tntgttgnag ggggtaccaa atttctttat 60
ttaaaggaat ggtacaaatc aaaaaactta atttaatttt tnggtacaac ttatagaaaa 120
ggttaaggaa accccaacat gcatgcactg ccttggtaac cagggnattc ccccncggct 180
ntggggaaat tagcccaang ctnagctttc attatcactn tcccccaggg tntgcttttc 240
aaaaaaattt nccgccnagc cnaatccggg cnctcccatc tggcgcaant tggtcacttg 300
gtcccccnat tctttaangg cttncacctn ctcattcggg tnatgtgtct caattaaatc 360
ccacngatgg ggtcattttt tntcnnttag ccagtttgtg nagttccgtt attganaaaa 420
ccan 424

<210> 12
<211> 426
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(426)
<223> n = A,T,C or G

<400> 12
tttttttttt ttttncttaa aagcttttat ctcctgctta cattacccat ctgttcttgc 60
atgttgtctg ctttttccac tagagccctt aacaacttaa tcatggttat tttaagggct 120
ctaataattc cnaaactggt atcataaata agtctcgttc tnatgcttgt ttctctcta 180

```
tcacactgtg ttngttgctt tttnacatgc tttgtaattt ttggctgaaa gctgaaaaat    240
nacatacctg gttntacaac ctgaggtaan cagccttnta gtgtgaggtt ttatatntta    300
ctggctaaga gctnggcnct gttnantant tgttgtanct ntatatgcca naggctttna    360
tttccnctng tgtccttgct tnagtacccc attnttttag gggttcccta naaactctat    420
ctnaat                                                               426
```

<210> 13
<211> 419
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(419)
<223> n = A,T,C or G

<400> 13

```
tttttttttt tttttnagat agactctcac tctttcgccc aggctggagt gcagtggcgc     60
aatcaaggct cactgcaacc tctgccttat aaagcatttn ctaaaggtac aagctaaatt    120
ttaaaaatat ctctncacaa ctaatgtata acaaaaatta gttctacctc ataaacncnt    180
ggctcagccc tcgnaacaca tttccctgtt ctcaactgat gaacactcca naaacagaac    240
anatntaagc ttttccaggc ccagaaaagc tcgcgagggg atttgctntg tgtgtgacac    300
acttgccacc ctgtggcagc acagctccac acntgctttg ggccgcattt gcaagttctc    360
tgtaanccccc ctgnaagacc cggatcagct gggtngaaat tgcangcnct cttttggca    419
```

<210> 14
<211> 400
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(400)
<223> n = A,T,C or G

<400> 14

```
aanccattgc caagggtatc cggaggattg tggctgtcac aggtnccgag gcccanaagg     60
ccctcaggaa agcaaagagc ttgaaaaatg tctctctgtc atggaagccn aagtgaaggc    120
tcanactgct ccaacaagga tntgcanagg gagatcgcta accttggaga ggccctggcc    180
actgcagtcn tcccccantg gcagaaggat gaattgcggg agactctcan atcccttang    240
gaaggtcgtg gatnacttgg accgagcctc nnaagccaat ntccagaaca agtgttggag    300
aagacaaagc anttcatcga cgccaacccc naccggcctc tnttctcctg ganattgana    360
gcggcgcccc cgcccagggc cttaataanc cntgaagctn                          400
```

<210> 15
<211> 395
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(395)
<223> n = A,T,C or G

<400> 15

```
tgctttgctg cgtccaggaa gattagatng aanaatacat attgatttgc caaatgaaca     60
agcgagatta gacntactga anatccatgc aggtcccatt acaaagcatg gtgaaataga    120
tgatgaagca attgtgaagc tatcggatgg ctttnatgga gcagatctga gaaatgtttg    180
tactgaagca ggtatgttcg caattcgtgc tgatcatgat tttgtagtac aggaagactt    240
catgaaagcn gtcagaanag tggctnattc tnaaagctgg agtctaaatt ggacnacnac    300
ctntgtattt actgttggan ttttgatgct gcatgacaga ttttgcttan tgtaaaaatn    360
aagttcaaga aaattatgtt agttttggcc attat                               395
```

<210> 16
<211> 404
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(404)
<223> n = A,T,C or G

<400> 16
```
ccaccactaa aatcctggct gagccctacn agtacctgtg cccctccccc aggacgagat      60
nagggcacac cctttaagtn aggtgacagg tcacctttaa gtgaggacag tcagctnaat     120
ttcacctctt gggcttgagt acctggttct cgtgccctga ggcgacnctn agccctgcag     180
ctnccatgta cgtgctgcca atngtcttga tcttctccac gccnctnaac ttgggcttca     240
gtaggagctg caggcnagaa ngaagcggtt aacagcgcca ctccatagcc gcagccnggc     300
tgcccctgct tctcaaggag gggtgtgggg ttcctccacc atcgccgccc ttgcaaacac     360
ntctcanggc ttccctnccg gctnancgca ngacttaagc atgg                      404
```

<210> 17
<211> 360
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(360)
<223> n = A,T,C or G

<400> 17
```
ggccagaagc tttccacaaa ccagtgaagg tggcagcaaa gaaagcctct tagacnagga      60
gctggcagca gctgctatct ngatngacng cagaaaccaa ccactaattc agcaaacaca     120
acctcatacc tnaccgcttc cctttnaatg gccttcggtg tgtgcgcaca tgggcacgtg     180
cggggagaac catacttatt cccctnttcc cggcctacca cctctnctcc cccttctctt     240
ctctncaatt actntctccn ctgctttntt ctnancacta ctgctngtnt cnanagccng     300
cccgcaatta cctggcaaaa ctcgcgaccc ttcgggcagc gctaaanaat gcacatttac     360
```

<210> 18
<211> 316
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(316)
<223> n = A,T,C or G

<400> 18
```
atacatatac acatatatga ttttagatag agccatatac ctngaagtag tanatttgtt      60
tgtgtgtata tgtatgtgtc tactcatttt aaataaactt gtgatagaga tgtaattntg     120
agccagtttt tcatttgctt aaatnactca ccaagtaact aattaagttn tctttactct     180
taatgttnag tagtgagatt ctgttgaagg tgatattaaa aaccattcta tattaattaa     240
cattcatgtt gttttttaaa agcttatttg aaatcnaatt atgattattt ttcataccag     300
tcgatnttat gtangt                                                     316
```

<210> 19
<211> 350
<212> DNA
<213> Homo sapien

<220>

<221> misc_feature
<222> (1)...(350)
<223> n = A,T,C or G

<400> 19
```
aagggatgca nataatgctg tgtatgagct tgatggaaaa gaactctgta gtgaaagggt      60
tactattgaa catgctnggg ctcggtcacg aggtggaaga ggtagaggac gatactctga     120
ccgttttagt agtcgcagac ctcgaaatga tagacgaaat gctccacctg taagaacaga     180
anatcgtctt atagttgaga atttatcctc aagagtcagc tggcaggttt gttganatac     240
agttttgagt tnttttgatg tggcttttta aaaaagttat gggttactna tgttatattg     300
ttttattaaa agtagttttn aattaatgga tntgatggaa ttgttgtttt                350
```

<210> 20
<211> 367
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(367)
<223> n = A,T,C or G

<400> 20
```
gntnnncnca agatcctnct ntcccccngg gcngccccnc cnccngtnat naccggtttn      60
ntaanatcnn gccgcncccg aagtctcnct nntgccgaga tgnccctat ncncnnatgn      120
ncaattntga cctnnggcga anaatggcng nngtgtatca gtntccnctc tgnggnctct     180
tagnatctga ccactangac ccnctatcct ctcaaaccct gtannc ngcc ctaatttgtg    240
ccaattagtg catgntanag cntcctggcc cagatggcnt ccatatcctg gtncggcttc     300
cgcccctacc angncatccn catctactag agcttatccg ctncntgngg cgcaccggnt     360
ccccnct                                                               367
```

<210> 21
<211> 366
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(366)
<223> n = A,T,C or G

<400> 21
```
cccaacacaa tggtctaagt anaactgtat tgctctgtag tatagttcca cattggcaac      60
ctacaatggg aaaatccata cataagtcag ttacttcctn atgagctttc tccttctgaa     120
tcctttatct tctgaagaaa gtacacacct tggtnatgat atctttgaat tgcccttctt     180
tccaggcatc agttggatga ttcatcatgg taattatggc attatcatat tcttcatact     240
tgtcatacga aaacaccagt tctgcccnna gatgagcttg ttctgcagct cttagcacct     300
tgggaatatt cactctagac cagaaacagc tcccggtgct ccctcatttt ctgaggctta     360
aatttn                                                                366
```

<210> 22
<211> 315
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(315)
<223> n = A,T,C or G

<400> 22
```
acttaatgca atctctggag gataatttgg atcaagaaat aaagaanaaa tgaattagga      60
```

```
gaagaaatna ctgggtnata tttcaatatt ttagaacttt aanaatgttg actatgattt     120
caatatattt gtnaaaactg agatacangt ttgacctata tctgcatttt gataattaaa     180
cnaatnnatt ctatttnaat gttgtttcag agtcacagca cagactgaaa cttttttga      240
atacctnaat atcacacttn tncttnnaat gatgttgaag acaatgatga catgccttna     300
gcatataatg tcgac                                                      315
```

<210> 23
<211> 202
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(202)
<223> n = A,T,C or G

<400> 23

```
actaatccag tgtggtgnaa ttccattgtg ttgggcaact caggatatta aatttatnat      60
ttaaaaattc ccaagagaaa naaactccag gccctgattg tttcactggg gaattttacc     120
aaatgttnca nnaaganatg acgctgattc tgtnaaatct ttttcagaag atagaggaga     180
acacccaccg nttcatttta tg                                              202
```

<210> 24
<211> 365
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(365)
<223> n = A,T,C or G

<400> 24

```
ggatttcttg cccttttctc ccttttaag tatcaatgta tgaaatccac ctgtaccacc       60
ctttctgcca tacaaccgct accacatctg gctcctagaa cctgttttgc tttcatagat     120
ggatctcgga accnagtgtt nacttcattt ttaaacccca ttttagcaga tngtttgctn     180
tggtctgtct gtattcacca tggggcctgt acacaccacg tgtggttata gtcaaacaca     240
gtgccctcca ttgtggccac atgggagacc catnacccna tactgcatcc tgggctgatn     300
acggcactgc atctnacccg acntgggatt gaacccgggg tgggcagcng aattgaacag     360
gatca                                                                 365
```

<210> 25
<211> 359
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(359)
<223> n = A,T,C or G

<400> 25

```
gtttcctgct tcaacagtgc ttggacggaa cccggcgctc gttccccacc ccggccggcc      60
gcccatagcc agccctccgt cacctcttca ccgcaccctc ggactgcccc aaggcccccg     120
ccgccnctcc ngcgccncgc agccaccgcc gccnccncca cctctccttn gtcccgccnt     180
nacaacgcgt ccacctcgca ngttcgccng aactaccacc nggactcata ngccgccctc     240
aaccgcccga tcaacctgga gctctncccc ccgacnttaa cctttccntg tcttacttac     300
nttaaccgcc gnttattttg cttnaaaaga acttttcccc aatactttct ttcaccnnt      359
```

<210> 26
<211> 400
<212> DNA

<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(400)
<223> n = A,T,C or G

<400> 26
agtgaaacag tatatgtgaa aaggagtttg tgannagcta cataaaaata ttagatatct 60
ttataatttc caataggata ctcatcagtt ttgaataana gacatattct agagaaacca 120
ggtttctggt ttcagatttg aactctcaag agcttggaag ttatcactcc catcctcacg 180
acnacnaana aatctnaacn aacngaanac caatgacttt tcttagatct gtcaaagaac 240
ttcagccacg aggaaaacta tcnccctnaa tactggggac tggaaagaga gggtacagag 300
aatcacagtg aatcatagcc caagatcagc ttgcccggag ctnaagctng tacgatnatt 360
acttacaggg accacttcac agtnngtnga tnaantgccn 400

<210> 27
<211> 366
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(366)
<223> n = A,T,C or G

<400> 27
gaatttctta gaaactgaag tttactctgt tccaagatat atcttcactg tcttaatcaa 60
agggcgctng aatcatagca aatattctca tctttcaact aactttaagt agttntcctg 120
gaattttaca ttttccagaa aacactcctt tctgtatctg tgaaagaaag tgtgcctcag 180
gctgtagact gggctgcact ggacacctgc ggggactct ggctnagtgn ggacatggtc 240
agtattgatt ttcctcanac tcagcctgtg tagctntgaa agcatggaac agattacact 300
gcagttnacg tcatcccaca catcttggac tccnagaccc ggggaggtca catagtccgt 360
tatgna 366

<210> 28
<211> 402
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(402)
<223> n = A,T,C or G

<400> 28
agtgggagcc tcctccttcc ccactcagtt ctttacatcc ccgaggcgca gctgggcnaa 60
ggaagtggcc agctgcagcg cctcctgcag gcagccaacg ttcttgcctg tggcctgtgc 120
agacacatcc ttgccaccac ctttaccgtc catcangcct gacacctgct gcacccactc 180
gctngctttt aagccccgat nggctgcatt ctgggggact tgacacaggc ncgtgatctt 240
gccagcctca ttgtccaccg tgaagagcat ggcaaaaagt ctgaggggag tgcatcttga 300
anagcttcaa ggcttcattc agggccttng ctnaggcgcc nctctccatc tccnggaata 360
acnagaggct ggtnngggtn actntcaata aactgcttcg tc 402

<210> 29
<211> 175
<212> DNA
<213> Homo sapien

<400> 29
cggacgggca tgaccggtcc ggtcagctgg gtggccagtt tcagttcttc agcagaactg 60
tctcccttct tgggggccga gggcttcctg gggaagagga tgagtttgga gcggtactcc 120

ttcagccgct gcacgttggt ctgcagggac tccgtggact tgttccgcct cctcg 175

<210> 30
<211> 360
<212> DNA
<213> Homo sapien

<400> 30
ttgtatttct tatgatctct gatgggttct tctcgaaaat gccaagtgga agactttgtg 60
gcatgctcca gatttaaatc cagctgaggc tccctttgtt ttcagttcca tgtaacaatc 120
tggaaggaaa cttcacggac aggaagactg ctggagaaga gaagcgtgtt agcccatttg 180
aggtctgggg aatcatgtaa agggtaccca gacctcactt ttagttattt acatcaatga 240
gttctttcag ggaaccaaac ccagaattcg gtgcaaaagc caaacatctt ggtgggattt 300
gataaatgcc ttgggacctg gagtgctggg cttgtgcaca ggaagagcac cagccgctga 360

<210> 31
<211> 380
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(380)
<223> n = A,T,C or G

<400> 31
acgctctaag cctgtccacg agctcaatag ggaagcctgt gatgactaca gactttgcga 60
acgctacgcc atggtttatg gatacaatgc tgcctataan cgctacttca ggaagcgccg 120
agggaccnaa tgagactgag ggaagaaaaa aaatctcttt ttttctggag gctggcacct 180
gattttgtat ccccctgtnn cagcattncn gaaatacata ggcttatata caatgcttct 240
ttcctgtata ttctcttgtc tggctgcacc ccttnttccc gcccccagat tgataagtaa 300
tgaaagtgca ctgcagtnag ggtcaangga gactcancat atgtgattgt tccntnataa 360
acttctggtg tgatactttc 380

<210> 32
<211> 440
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(440)
<223> n = A,T,C or G

<400> 32
gtgtatggga gcccctgact cctcacgtgc ctgatctgtg cccttggtcc caggtcaggc 60
ccaccccctg cacctccacc tgccccagcc cctgcctctg ccccaagtgg ggccagctgc 120
cctcacttct ggggtggatg atgtgacctt cctngggggа ctgcggaagg gacaagggtt 180
ccctgaagtc ttacggtcca acatcaggac caagtcccat ggacatgctg acagggtccc 240
caggggagac cgtntcanta gggatgtgtg cctggctgtg tacgtgggtg tgcagtgcac 300
gtganaagca cgtggcggct tctgggggcc atgtttgggg aaggaagtgt gcccnccacc 360
cttggagaac ctcagtcccn gtagccccct gccctggcac agcngcatnc acttcaaggg 420
caccctttgg gggttggggt 440

<210> 33
<211> 345
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(345)

<223> n = A,T,C or G

<400> 33
tattttaaca atgtttatta ttcatttatc cctctataga accaccaccc acaccgagga 60
gattatttgg agtgggtccc aacctagggc ctggactctg aaatctaact ccccacttcc 120
ctcattttgt gacttaggtg ggggcatggt tcagtcagaa ctggtgtctc ctattggatc 180
gtgcagaagg aggacctagg cacacacata tggtggccac acccaggagg gttgattggc 240
aggctggaag acaaaagtct cccaataaag gcacttttac ctcaaagang gggtgggagt 300
tggtctgctg ggaatgttgt tgttggggtg gggaagantt atttc 345

<210> 34
<211> 440
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(440)
<223> n = A,T,C or G

<400> 34
tgtaattttt ttattggaaa acaaatatac aacttggaat ggattttgag gcaaattgtg 60
ccataagcag attttaagtg gctaaacaaa gtttaaaaag caagtaacaa taaaagaaaa 120
tgtttctggt acaggaccag cagtacaaaa aaatagtgta cgagtacctg gataatacac 180
ccgttttgca atagtgcaac ttttaagtac atattgttga ctgtccatag tccacgcaga 240
gttacaactc cacacttcaa caacaacatg ctgacagttc ctaaagaaaa ctactttaaa 300
aaaggcataa cccagatgtt ccctcatttg accaactcca tctnagttta gatgtgcaga 360
agggcttana ttttcccaga gtaagccnca tgcaacatgt tacttgatca attttctaaa 420
ataaggtttt aggacaatga 440

<210> 35
<211> 540
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(540)
<223> n = A,T,C or G

<400> 35
atagatggaa tttattaagc ttttcacatg tgatagcaca tagttttaat tgcatccaaa 60
gtactaacaa aaactctagc aatcaagaat ggcagcatgt tattttataa caatcaacac 120
ctgtggcttt taaaatttgg ttttcataag ataatttata ctgaagtaaa tctagccatg 180
cttttaaaaa atgctttagg tcactccaag cttggcagtt aacatttggc ataaacaata 240
ataaaacaat cacaatttaa taaataacaa atacaacatt gtaggccata atcatataca 300
gtataaggga aaaggtggta gtgttganta agcagttatt agaatagaat accttggcct 360
ctatgcaaat atgtctagac actttgattc actcagccct gacattcagt tttcaaagtt 420
aggaaacagg ttctacagta tcattttaca gtttccaaca cattgaaaac aagtagaaaa 480
tgatganttg atttttatta atgcattaca tcctcaagan ttatcaccaa cccctcaggt 540

<210> 36
<211> 555
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(555)
<223> n = A,T,C or G

<400> 36

```
cttcgtgtgc ttgaaaattg gagcctgccc ctcggcccat aagcccttgt tgggaactga      60
gaagtgtata tggggcccaa nctactggtg ccagaacaca gagacagcag cccantgcaa     120
tgctgtcgag cattgcaaac gccatgtgtg gaactaggag gaggaatatt ccatcttggc     180
agaaaccaca gcattggttt ttttctactt gtgtgtctgg gggaatgaac gcacagatct     240
gtttgacttt gttataaaaa tagggctccc ccacctcccc cntttctgtg tnctttattg     300
tagcantgct gtctgcaagg gagcccctan cccctggcag acananctgc ttcagtgccc     360
ctttcctctc tgctaaatgg atgttgatgc actggaggtc ttttancctg cccttgcatg     420
gcncctgctg gaggaagana aaactctgct ggcatgaccc acagtttctt gactggangc     480
cntcaaccct cttggttgaa gccttgttct gaccctgaca tntgcttggg cnctgggtng     540
gnctgggctt ctnaa                                                      555
```

<210> 37
<211> 280
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(280)
<223> n = A,T,C or G

<400> 37

```
ccaccgacta taagaactat gccctcgtgt attcctgtac ctgcatcatc caactttttc      60
acgtggattt tgcttggatc ttggcaagaa accctaatct ccctccagaa acagtggact     120
ctctaaaaaa tatcctgact tctaataaca ttgatntcaa gaaaatgacg gtcacagacc     180
aggtgaactg ccccnagctc tcgtaaccag gttctacagg gaggctgcac ccactccatg     240
ttncttctgc ttcgctttcc cctaccccac cccccgccat                           280
```

<210> 38
<211> 303
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(303)
<223> n = A,T,C or G

<400> 38

```
catcgagctg gttgtcttct tgcctgccct gtgtcgtaaa atgggggtcc cttactgcat      60
tatcaaggga aaggcaagac tgggacgtct agtccacagg aagacctgca ccactgtcgc     120
cttcacacag gtgaactcgg aagacaaagg cgctttggct nagctggtgn aagctatcag     180
gaccaattac aatgacngat acgatnagat ccgccntcac tggggtagca atgtcctggg     240
tcctaagtct gtggctcgta tcgccnagct cgaanaggcn aangctaaag aacttgccac     300
taa                                                                   303
```

<210> 39
<211> 300
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(300)
<223> n = A,T,C or G

<400> 39

```
gactcagcgg ctggtgctct tcctgtgcac aagcccagca ctccaggtcc caaggcattt      60
atcaaatccc accaagatnt ttggcttttg caccgaattc tgggtttggt tccctnaaag     120
aactcattga tgtaaatnac tnaaagtgag gtctgggtac cctttacatg attccccaga     180
cctcanatgg gctaacacgc ttctcttctc cagcagtctt cctntccgtg aagttacctt     240
ccagattgtt acatggaact gaanacaaag ggagcctcag ctngatttaa atctggagca     300
```

<210> 40
<211> 318
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(318)
<223> n = A,T,C or G

<400> 40
```
cccaacacaa tggctgagga caaatcagtt ctctgtgacc agacatgaga aggttgccaa      60
tgggctgttg ggcgaccaag gccttcccgg agtcttcgtc ctctatgagc tctcgcccat     120
gatggtgaag ctgacggaga agcacaggtc cttcacccac ttcctgacag gtgtgtgcgc     180
catcattggg ggcatgttca cagtggctgg actcatcgat tcgctcatct accactcagc     240
acgagccatc cagaaaaaaa ttgatctngg gaagacnacg tagtcaccct cggtncttcc     300
tctgtctcct ctttctcc                                                  318
```

<210> 41
<211> 302
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(302)
<223> n = A,T,C or G

<400> 41
```
acttagatgg ggtccgttca ggggatacca gcgttcacat ttttcctttt aagaaagggt      60
cttggcctga atgttcccca tccggacaca ggctgcatgt ctctgtnagt gtcaaagctg     120
ccatnaccat ctcggtaacc tactcttact ccacaatgtc tatnttcact gcagggctct     180
ataatnagtc cataatgtaa atgcctggcc caagacntat ggcctgagtt tatccnaggc     240
ccaaacnatt accagacatt cctcttanat tgaaaacgga tntctttccc ttggcaaaga     300
tc                                                                   302
```

<210> 42
<211> 299
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(299)
<223> n = A,T,C or G

<400> 42
```
cttaataagt ttaaggccaa ggcccgttcc attcttctag caactgacgt tgccagccga      60
ggtttggaca tacctcatgt aaatgtggtt gtcaactttg acattcctac ccattccaag     120
gattacatcc atcgagtagg tcgaacagct agagctgggc gctccggaaa ggctattact     180
tttgtcacac agtatgatgt ggaactcttc cagcgcatag aacacttnat tgggaagaaa     240
ctaccaggtt ttccaacaca ggatgatgag gttatgatgc tnacggaacg cgtcgctna      299
```

<210> 43
<211> 305
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(305)

<223> n = A,T,C or G

<400> 43
ccaacaatgt caagacagcc gtctgtgaca tcccacctcg tggcctcaan atggcagtca 60
ccttcattgg caatagcaca gccntccggg agctcttcaa gcgcatctcg gagcagttca 120
ctgccatgtt ccgccggaag gccttcctcc actggtacac aggcgagggc atggacaaga 180
tggagttcac cgaggctgag agcaacatga acgacctcgt ctctnagtat cagcagtacc 240
gggatgccac cgcagaaana ggaggaggat ttcggtnagg aggccgaaga aggaggcctg 300
aggca 305

<210> 44
<211> 399
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(399)
<223> n = A,T,C or G

<400> 44
tttctgtggg ggaaacctga tctcgacnaa attagagaat tttgtcagcg gtatttcggc 60
tggaacagaa cgaaaacnga tnaatctctg tttcctgtat taaagcaact cgatncccag 120
cagacacagc tccnaattga ttccttcttt ngattagcac aacagggaga aagaanatgc 180
ttaacgtatt aagagccnga gactaaacag agctttgaca tgtatgctta ggaaagagaa 240
agaagcagcn gcccgcgnaa ttngaagcng tttctgttgc cntggganaaa gaatttgagc 300
ttctttatta ggccaacgaa aaaccccgaa ananaggcnt tacnatacct tngaaaantc 360
tccngccnna aaaagaaaga agctttcnga ttcttaacc 399

<210> 45
<211> 440
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(440)
<223> n = A,T,C or G

<400> 45
gcgggagcag aagctaaagc caaagcccaa gagagtggca gtgccagcac tggtgccagt 60
accagtacca ataacagtgc cagtgccagt gccagcacca gtggtggctt cagtgctggt 120
gccagcctga ccgccactct cacatttggg ctcttcgctg gccttggtgg agctggtgcc 180
agcaccagtg gcagctctgg tgcctgtggt ttctcctaca agtgagattt taggtatctg 240
ccttggtttc agtggggaca tctggggctt angggcngg gataaggagc tggatgattc 300
taggaaggcc cangttggag aangatgtgn anagtgtgcc aagacactgc ttttggcatt 360
ttattccttt ctgtttgctg gangtcaatt gacccttnna ntttctctta cttgtgtttt 420
canatatngt taatcctgcc 440

<210> 46
<211> 472
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(472)
<223> n = A,T,C or G

<400> 46
gctctgtaat ttcacatttt aaaccttccc ttgacctcac attcctcttc ggccacctct 60
gtttctctgt tcctcttcac agcaaaaact gttcaaaaga gttgttgatt actttcattt 120

ccactttctc acccccattc tcccctcaat taactctcct tcatccccat gatgccatta 180
tgtggctntt attanagtca ccaaccttat tctccaaaac anaagcaaca aggactttga 240
cttctcagca gcactcagct ctggtncttg aaacaccccc gttacttgct attcctccta 300
cctcataaca atctccttcc cagcctctac tgctgccttc tctgagttct tcccagggtc 360
ctaggctcag atgtagtgta gctcaaccct gctacacaaa gnaatctcct gaaagcctgt 420
aaaaatgtcc atncntgtcc tgtgagtgat ctnccangna naataacaaa tt 472

<210> 47
<211> 550
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(550)
<223> n = A,T,C or G

<400> 47
ccttcctccg cctggccatc cccagcatgc tcatgctgtg catggagtgg tgggcctatg 60
aggtcgggag cttcctcagt ggtctgtatg aggatggatg acggggactg gtgggaacct 120
gggggccctg tctgggtgca aggcgacagc tgtctttctt caccaggcat cctcggcatg 180
gtggagctgg gcgctcagtc catcgtgtat gaactggcca tcattgtgta catggtccct 240
gcaggcttca gtgtggctgc cagtgtccgg gtangaaacg ctctgggtgc tggagacatg 300
gaagcaggca cggaagtcct ctaccgtttc cctgctgatt acagtgctct ttgctgtanc 360
cttcagtgtc ctgctgttaa gctgtaagga tcacntgggg tacattttta ctaccgaccg 420
agaacatcat taatctggtg gctcaggtgg ttccaattta tgctgtttcc cacctctttg 480
aagctcttgc tgctcaggta cacgccaatt ttgaaaagta aacaacgtgc ctcggagtgg 540
gaattctgct 550

<210> 48
<211> 214
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(214)
<223> n = A,T,C or G

<400> 48
agaaggacat aaacaagctg aacctgccca agacgtgtga tatcagcttc tcagatccag 60
acaacctcct caacttcaag ctggtcatct gtcctgatna gggcttctac nagagtggga 120
agtttgtgtt cagttttaag gtgggccagg gttacccgca tgatccccc aaggtgaagt 180
gtgagacnat ggtctatcac cccnacattg acct 214

<210> 49
<211> 267
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(267)
<223> n = A,T,C or G

<400> 49
atctgcctaa aatttattca aataatgaaa atnaatctgt tttaagaaat tcagtctttt 60
agtttttagg acaactatgc acaaatgtac gatggagaat tctttttgga tnaactctag 120
gtngaggaac ttaatccaac cggagctntt gtgaaggtca gaanacagga gagggaatct 180
tggcaaggaa tggagacnga gtttgcaaat tgcagctaga gtnaatngtt ntaaatggga 240
ctgctnttgt gtctcccang gaaagtt 267

<210> 50
<211> 300
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(300)
<223> n = A,T,C or G

<400> 50
gactgggtca aagctgcatg aaaccaggcc ctggcagcaa cctgggaatg gctggaggtg 60
ggagagaacc tgacttctct ttccctctcc ctcctccaac attactggaa ctctgtcctg 120
ttgggatctt ctgagcttgt ttccctgctg ggtgggacag aggacaaagg agaagggagg 180
gtctagaaga ggcagccctt ctttgtcctc tggggtnaat gagcttgacc tanagtagat 240
ggagagacca anagcctctg atttttaatt tccataanat gttcnaagta tatntntacc 300

<210> 51
<211> 300
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(300)
<223> n = A,T,C or G

<400> 51
gggtaaaatc ctgcagcacc cactctggaa aatactgctc ttaattttcc tgaaggtggc 60
cccctatttc tagttggtcc aggattaggg atgtggggta tagggcattt aaatcctctc 120
aagcgctctc caagcacccc cggcctgggg gtnagtttct catcccgcta ctgctgctgg 180
gatcaggttn aataaatgga actcttcctg tctggcctcc aaagcagcct aaaaactgag 240
gggctctgtt agaggggacc tccaccctnn ggaagtccga ggggctnggg aagggtttct 300

<210> 52
<211> 267
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(267)
<223> n = A,T,C or G

<400> 52
aaaatcaact tcntgcatta atanacanat tctanancag gaagtgaana taattttctg 60
cacctatcaa ggaacnnact tgattgcctc tattnaacan atatatcgag ttnctatact 120
tacctgaata ccnccgcata actctcaacc nanatncntc nccatgacac tcnttcttna 180
atgctantcc cgaattcttc attatatcng tgatgttcgn cctgntnata tatcagcaag 240
gtatgtnccn taactgccga nncaang 267

<210> 53
<211> 401
<212> DNA
<213> Homo sapien

<400> 53
agsctttagc atcatgtaga agcaaactgc acctatggct gagataggtg caatgaccta 60
caagattttg tgttttctag ctgtccagga aaagccatct tcagtcttgc tgacagtcaa 120
agagcaagtg aaaccatttc cagcctaaac tacataaaag cagccgaacc aatgattaaa 180
gacctctaag gctccataat catcattaaa tatgcccaaa ctcattgtga ctttttattt 240
tatatacagg attaaaatca acattaaatc atcttattta catggccatc ggtgctgaaa 300

```
ttgagcattt taaatagtac agtaggctgg tatacattag gaaatggact gcactggagg     360
caaatagaaa actaaagaaa ttagataggc tggaaatgct t                         401
```

<210> 54
<211> 401
<212> DNA
<213> Homo sapien

<400> 54

```
cccaacacaa tggataaaaa cacttatagt aaatggggac attcactata atgatctaag      60
aagctacaga ttgtcatagt tgttttcctg ctttacaaaa ttgctccaga tctggaatgc     120
cagtttgacc tttgtcttct ataatatttc ctttttttcc cctctttgaa tctctgtata     180
tttgattctt aactaaaatt gttctcttaa atattctgaa tcctggtaat taaaagtttg     240
ggtgtatttt ctttacctcc aaggaaagaa ctactagcta caaaaaatat tttggaataa     300
gcattgtttt ggtataaggt acatattttg gttgaagaca ccagactgaa gtaaacagct     360
gtgcatccaa tttattatag ttttgtaagt aacaatatgt a                         401
```

<210> 55
<211> 933
<212> DNA
<213> Homo sapien

<400> 55

```
tttactgctt ggcaaagtac cctgagcatc agcagagatg ccgagatgaa atcagggaac      60
tcctagggga tgggtcttct attacctggg aacacctgag ccagatgcct tacaccacga     120
tgtgcatcaa ggaatgcctc cgcctctacg caccggtagt aaactatccc ggttactcga     180
caaacccatc acctttccag atggacgctc cttacctgca ggaataactg tgtttatcaa     240
tatttgggct cttcaccaca acccctattt ctgggaagac cctcaggtct ttaacccctt     300
gagattctcc agggaaaatt ctgaaaaaat acatccctat gccttcatac cattctcagc     360
tggattaagg aactgcattg ggcagcattt tgccataatt gagtgtaaag tggcagtggc     420
attaactctg ctccgcttca agctggctcc agaccactca aggccaccca gctgtcgtca     480
agttgcctca agtccaagaa tggaatccat gtgtttgcaa aaaaagtttg ctaattttaa     540
gtccttttcg tataagaatt aakgagacaa ttttcctacc aaaggaagaa caaaaggata     600
aatataatac aaaatatatg tatatggttg tttgacaaat tatataactt aggatacttc     660
tgactggttt tgacatccat taacagtaat tttaatttct ttgctgtatc tggtgaaacc     720
cacaaaaaca cctgaaaaaa ctcaagctga gttccaatgc gaagggaaat gattggtttg     780
ggtaactagt ggtagagtgg ctttcaagca tagtttgatc aaaactccac tcagtatctg     840
cattactttt atctctgcaa atatctgcat gatagcttta ttctcagtta tctttcccca     900
taataaaaaa tatctgccaa aaaaaaaaaa aaa                                  933
```

<210> 56
<211> 480
<212> DNA
<213> Homo sapien

<400> 56

```
ggctttgaag catttttgtc tgtgctccct gatcttcagg tcaccaccat gaagttctta      60
gcagtcctgg tactcttggg agtttccatc tttctggtct ctgcccagaa tccgacaaca     120
gctgctccag ctgacacgta tccagctact ggtcctgctg atgatgaagc ccctgatgct     180
gaaaccactg ctgctgcaac cactgcgacc actgctgctc ctaccactgc aaccaccgct     240
gcttctacca ctgctcgtaa agacattcca gttttaccca aatgggttgg ggatctcccg     300
aatggtagag tgtgtccctg agatggaatc agcttgagtc ttctgcaatt ggtcacaact     360
attcatgctt cctgtgattt catccaacta cttaccttgc ctacgatatc ccctttatct     420
ctaatcagtt tattttcttt caaataaaaa ataactatga gcaacaaaaa aaaaaaaaaa     480
```

<210> 57
<211> 798
<212> DNA
<213> Homo sapien

<400> 57

```
agcctacctg gaaagccaac cagtcctcat aatggacaag atccaccagc tcctcctgtg      60
```

gactaacttt gtgatatggg aagtgaaaat agttaacacc ttgcacgacc aaacgaacga 120
agatgaccag agtactctta acccсttaga actgtttttc cttttgtatc tgcaatatgg 180
gatggtattg ttttcatgag cttctagaaa tttcacttgc aagtttattt ttgcttcctg 240
tgttactgcc attcctattt acagtatatt tgagtgaatg attatatttt taaaaagtta 300
catggggctt ttttggttgt cctaaactta caaacattcc actcattctg tttgtaactg 360
tgattataat ttttgtgata atttctggcc tgattgaagg aaatttgaga ggtctgcatt 420
tatatatttt aaatagattt gataggtttt taaattgctt tttttcataa ggtatttata 480
aagttatttg gggttgtctg ggattgtgtg aaagaaaatt agaaccccgc tgtatttaca 540
tttaccttgg tagtttattt gtggatggca gttttctgta gttttgggga ctgtggtagc 600
tcttggattg ttttgcaaat tacagctgaa atctgtgtca tggattaaac tggcttatgt 660
ggctagaata ggaagagaga aaaaatgaaa tggttgttta ctaattttat actcccatta 720
aaaattttta atgttaagaa aaccttaaat aaacatgatt gatcaatatg gaaaaaaaaa 780
aaaaaaaaaa aaaaaaa 798

<210> 58
<211> 280
<212> DNA
<213> Homo sapien

<400> 58
ggggcagctc ctgaccctcc acagccacct ggtcagccac cagctggggc aacgagggtg 60
gaggtcccac tgagcctctc gcctgccccc gccactcgtc tggtgcttgt tgatccaagt 120
cccctgcctg gtcccccaca aggactccca tccaggcccc ctctgccctg ccccttgtca 180
tggaccatgg tcgtgaggaa gggctcatgc cccttattta tgggaaccat ttcattctaa 240
cagaataaac cgagaaggaa accagaaaaa aaaaaaaaaa 280

<210> 59
<211> 382
<212> DNA
<213> Homo sapien

<400> 59
aggcgggagc agaagctaaa gccaaagccc aagagagtgg cagtgccagc actggtgcca 60
gtaccagtac caataacagt gccagtgcca gtgccagcac cagtggtggc ttcagtgctg 120
gtgccagcct gaccgccact ctcacatttg ggctcttcgc tggccttggt ggagctggtg 180
ccagcaccag tggcagctct ggtgcctgtg gtttctccta caagtgagat tttagatatt 240
gttaatcctg ccagtctttc tcttcaagcc agggtgcatc ctcagaaacc tactcaacac 300
agcactctag gcagccacta tcaatcaatt gaagttgaca ctctgcatta aatctatttg 360
ccattaaaaa aaaaaaaaaa aa 382

<210> 60
<211> 602
<212> DNA
<213> Homo sapien

<400> 60
tgaagagccg cgcggtggag ctgctgcccg atgggactgc caaccttgcc aagctgcagc 60
ttgtggtgga gaatagtgcc cagcgggtca tccacttggc gggtcagtgg gagaagcacc 120
gggtcccatc ctcgtgagta ccgccactcc gaaagctgca ggattgcaga gagctggaat 180
cttctcgacg gctggcagag atccaagaac tgcaccagag tgtccgggcg gctgctgaag 240
aggcccgcag gaaggaggag gtctataagc agctgatgtc agagctggag actctgccca 300
gagatgtgtc ccggctggcc tacacccagc gcatcctgga gatcgtgggc aacatccgga 360
agcagaagga agagatcacc aagatcttgt ctgatacgaa ggagcttcag aaggaaatca 420
actccctatc tgggaagctg gaccggacgt ttgcggtgac tgatgagctt gtgttcaagg 480
atgccaagaa ggacgatgct gttcggaagg cctataagta tctagctgct ctgcacgaga 540
actgcagcca gctcatccag accatcgagg acacaggcac catcatgcgg gaggttcgag 600
ac 602

<210> 61
<211> 1368
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(1368)
<223> n = A,T,C or G

<400> 61
```
ccagtgagcg cgcgtaatac gactcactat agggcgaatt gggtaccggg cccccccctcg      60
agcggccgcc cttttttttt ttttttttatt gatcagaatt caggctttat tattgagcaa     120
tgaaaacagc taaaacttaa ttccaagcat gtgtagttaa agtttgcaaa gtgggatatt     180
gttcacaaaa cacattcaat gtttaaacac tatttatttg aagaacaaaa tatatttaaa     240
attgtttgct tctaaaaagc ccatttccct ccaagtctaa actttgtaat ttgatattaa     300
gcaatgaagt tattttgtac aatctagtta aacaagcaga atagcactag gcagaataaa     360
aaattgcaca gacgtatgca attttccaag atagcattct ttaaattcag ttttcagctt     420
ccaaagattg gttgcccata atagacttaa acatataatg atggctaaaa aaaataagta     480
tacgaaaatg taaaaaagga aatgtaagtc cactctcaat ctcataaaag gtgagagtaa     540
ggatgctaaa gcaaaataaa tgtaggttct ttttttctgt ttccgtttat catgcaatct     600
gcttctttga tatgccttag ggttacccat ttaagttaga ggttgtaatg caatggtggg     660
aatgaaaatt gatcaaatat acaccttgtc atttcatttc aaattgcggg ctggaaactt     720
ccaaaaaaag ggtaggcatg aagaaaaaaa aaatcmaatc agaacctctt caggggtttg     780
kgktctgata tggcagacar gatacaagtc ccaccaggag atggagcaat tcaaaataag     840
ggtaatgggc tgacaaggta ttattgccag catgggacag aatgagcaac aggctgaaaa     900
gtttttggat tatatagcac ctagagtctc tgatgtaggg aatttttgtt agtcaaacat     960
acgctaaact tccaagggaa aatctttcag gtagcctaag cttgcttttc tagagtgatg    1020
agttgcattg ctactgtgat tttttgaaaa caaactgggt ttgtacaagt gagaaagact    1080
agagagaaag attttagtct gtttagcaga agccatttta tctgcgtgca catggatcaa    1140
tatttctgat cccctatacc ccaggaaggg caaaatccca aagaaatgtg ttagcaaaat    1200
tggctgatgc tatcatattg ctatggacat tgatcttgcc caacacaatg gaattccacc    1260
acactggact agtggatcca ctagttctag agcggccggc caccgcggtg gagctccagc    1320
ttttgttccc tttagtgagg gttaattgcg cgcttggcgt aatcatnn                 1368
```

<210> 62
<211> 924
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(924)
<223> n = A,T,C or G

<400> 62
```
caaaggnaca ggaacagctt gnaaagtact gncatncctn cctgcaggga ccagcccttt      60
gcctccaaaa gcaataggaa atttaaaaga tttncactga gaaggggncc acgtttnart     120
tntnaatgtn tcargnanar tnccttncaa atgncrnctn cactnactnr gnatttgggt     180
tnccgnrtnc mgnactatnt caggtttgaa aaactggatc tgccacttat cagttatgtg     240
accttaaaga actccgttaa tttctcagag cctcagtttc cttgtctata agttgggagt     300
aatattaata ctatcatttt tccaaggatt gatgtgaaca ttaatgaggt gaaatgacag     360
atgtgtatca tggttcctaa taaacatcca aaatatagta cttactattg tcattattat     420
tacttgtttg aagctaaaga cctcacaata gaatcccatc cagcccacca gacagagytc     480
tgagttttct agtttggaag agctattaaa taacaacktc tagtgtcaat tctatacttg     540
ttatggtcaa gtaactgggc tcagcatttt acattcattg tctctttaag ttctagcaat     600
gtgaagcagg aactatgatt atattgacta cataaatgaa gaaattgagg ctcagataca     660
ttaagtaatt ctcccagggt cacacagcta gaactggcaa agcctgggat tgatccatga     720
tcttccagca ttgaagaatc ataaatgtaa ataactgcaa ggccttttcc tcagaagagc     780
tcctggtgct tgcaccaacc cactagcact tgttctctac aggggaacat ctgtgggcct     840
gggaatcact gcacgtcgca agagatgttg cttctgatga attattgttc ctgtcagtgg     900
tgtgaaggca aaaaaaaaaa aaaa                                             924
```

<210> 63
<211> 1079
<212> DNA

<213> Homo sapien

<400> 63

```
agtcccaaga actcaataat ctcttatgtt ttcttttgaa gacttatttt aaatattaac      60
tatttcggtg cctgaatgga aaaatataaa cattagctca gagacaatgg ggtacctgtt     120
tggaatccag ctggcagcta taagcaccgt tgaaaactct gacaggcttt gtgccctttt     180
tattaaatgg cctcacatcc tgaatgcagg aatgtgttcg tttaaataaa cattaatctt     240
taatgttgaa ttctgaaaac acaaccataa atcatagttg gtttttctgt gacaatgatc     300
tagtacatta tttcctccac agcaaaccta cctttccaga aggtggaaat tgtatttgca     360
acaatcaggg caaaacccac acttgaaaag catttacaa tattatatct aagttgcaca      420
gaagacccca gtgatcacta ggaaatctac cacagtccag tttttctaat ccaagaaggt     480
caaacttcgg ggaataatgt gtccctcttc tgctgctgct ctgaaaaata ttcgatcaaa     540
acgaagttta caagcagcag ttattccaag attagagttc atttgtgtat cccatgtata     600
ctggcaatgt ttaggtttgc ccaaaaactc ccagacatcc acaatgttgt tgggtaaacc     660
accacatctg gtaacctctc gatcccttag atttgtatct cctgcaaata taactgtagc     720
tgactctgga gcctcttgca ttttctttaa aaccattttt aactgattca ttcgttccgc     780
agcatgccct ctggtgctct ccaaatggga tgtcataagg caaagctcat ttcctgacac     840
attcacatgc acacataaaa ggtttctcat catttggta cttggaaaag gaataatctc      900
ttggcttttt aatttcactc ttgatttctt caacattata gctgtgaaat atccttcttc     960
atgacctgta ataatctcat aattacttga tctcttcttt aggtagctat aatatggggg    1020
aataacttcc tgtagaaata tcacatctgg gctgtacaaa gctaagtagg aacacaccc     1079
```

<210> 64
<211> 1001
<212> DNA
<213> Homo sapien

<400> 64

```
gaatgtgcaa cgatcaagtc agggtatctg tggtatccac cactttgagc atttatcgat      60
tctatatgtc aggaacattt caagttatct gttctagcaa ggaaatataa aatacttata     120
gttaactatg gcctatctac agtgcaacta aaaactagat tttattcctt tccacctgtg     180
ggtttgtatt catttaccac cctcttttca ttccctttct cacccacaca ctgtgccggg     240
cctcaggcat atactattct actgtctgtc tctgtaagga ttatcatttt agcttccaca     300
tatgagagaa tgcatgcaaa gtttttcttt ccatgtctgg cttatttcac ttaacataat     360
gacctccgct tccatccatg ttatttatat tacccaatag tgttcataaa tatatataca     420
cacatatata ccacattgca tttgtccaat tattcattga cggaaactgg ttaatgttat     480
atcgttgcta ttgtggatag tgctgcaata aacacgcaag tggggatata atttgaagag     540
ttttttgtt gatgttcctc caaattttaa gattgttttg tctatgtttg tgaaaatggc      600
gttagtattt tcatagagat tgcattgaat ctgtagattg ctttgggtaa gtatggttat     660
tttgatggta ttaatttttt cattccatga agatgagatg tctttccatt gtttgtgtcc     720
tctacatttt ctttcatcaa agttttgttg tatttttgaa gtagatgtat ttcaccttat     780
agatcaagtg tattccctaa atattttatt tttgtagcta ttgtagatga aattgccttc     840
ttgatttctt tttcacttaa ttcattatta gtgtatggaa atgttatgga tttttatttg     900
ttggttttta atcaaaaact gtattaaact tagagttttt tgtggagttt ttaagttttt     960
ctagatataa gatcatgaca tctaccaaaa aaaaaaaaaa a                        1001
```

<210> 65
<211> 575
<212> DNA
<213> Homo sapien

<400> 65

```
acttgatata aaaaggatat ccataatgaa tattttatac tgcatccttt acattagcca      60
ctaaatacgt tattgcttga tgaagacctt tcacagaatc ctatggattg cagcatttca     120
cttggctact tcatacccat gccttaaaga ggggcagttt ctcaaaagca gaaacatgcc     180
gccagttctc aagttttcct cctaactcca tttgaatgta agggcagctg gcccccaatg     240
tggggaggtc cgaacatttt ctgaattccc attttcttgt tcgcggctaa atgacagttt     300
ctgtcattac ttagattccc gatctttccc aaaggtgttg atttacaaag aggccagcta     360
atagccagaa atcatgaccc tgaaagagag atgaaatttc aagctgtgag ccaggcagga     420
gctccagtat ggcaaaggtt cttgagaatc agccatttgg tacaaaaaag atttttaaag     480
cttttatgtt ataccatgga gccatagaaa ggctatggat tgtttaagaa ctattttaaa     540
gtgttccaga cccaaaaagg aaaaaaaaaa aaaaa                                575
```

<210> 66
<211> 831
<212> DNA
<213> Homo sapien

<400> 66
attgggctcc ttctgctaaa cagccacatt gaaatggttt aaaagcaagt cagatcaggt 60
gatttgtaaa attgtattta tctgtacatg tatgggcttt taattcccac caagaaagag 120
agaaattatc tttttagtta aaaccaaatt tcacttttca aaatatcttc caacttattt 180
attggttgtc actcaattgc ctatatatat atatatatat gtgtgtgtgt gtgtgtgcgc 240
gtgagcgcac gtgtgtgtat gcgtgcgcat gtgtgtgtat gtgtattatc agacataggt 300
ttctaacttt tagatagaag aggagcaaca tctatgccaa atactgtgca ttctacaatg 360
gtgctaatct cagacctaaa tgatactcca tttaatttaa aaaagagttt taaataatta 420
tctatgtgcc tgtatttccc ttttgagtgc tgcacaacat gttaacatat tagtgtaaaa 480
gcagatgaaa caaccacgtg ttctaaagtc tagggattgt gctataatcc ctatttagtt 540
caaaattaac cagaattctt ccatgtgaaa tggaccaaac tcatattatt gttatgtaaa 600
tacagagttt taatgcagta tgacatccca caggggaaaa gaatgtctgt agtgggtgac 660
tgttatcaaa tattttatag aatacaatga acggtgaaca gactggtaac ttgtttgagt 720
tcccatgaca gatttgagac ttgtcaatag caaatcattt ttgtatttaa atttttgtac 780
tgatttgaaa aacatcatta aatatcttta aaagtaaaaa aaaaaaaaaa a 831

<210> 67
<211> 590
<212> DNA
<213> Homo sapien

<400> 67
gtgctctgtg tattttttta ctgcattaga cattgaatag taatttgcgt taagatacgc 60
ttaaaggctc tttgtgacca tgtttccctt tgtagcaata aaatgttttt tacgaaaact 120
ttctccctgg attagcagtt taaatgaaac agagttcatc aatgaaatga gtatttaaaa 180
taaaaatttg ccttaatgta tcagttcagc tcacaagtat tttaagatga ttgagaagac 240
ttgaattaaa gaaaaaaaaa ttctcaatca tatttttaaa atataagact aaaattgttt 300
ttaaaacaca tttcaaatag aagtgagttt gaactgacct tatttatact ctttttaagt 360
ttgttccttt tccctgtgcc tgtgtcaaat cttcaagtct tgctgaaaat acatttgata 420
caaagttttc tgtagttgtg ttagttcttt tgtcatgtct gtttttggct gaagaaccaa 480
gaagcagact tttcttttaa aagaattatt tctctttcaa atatttctat ccttttttaaa 540
aaattccttt ttatggctta tatacctaca tatttaaaaa aaaaaaaaaa 590

<210> 68
<211> 291
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(291)
<223> n = A,T,C or G

<400> 68
gttccctttt ccggtcggcg tggtcttgcg agtggagtgt ccgctgtgcc cgggcctgca 60
ccatgagcgt cccggccttc atcgacatca gtgaagaaga tcaggctgct gagcttcgtg 120
cttatctgaa atctaaagga gctgagattt cagaagagaa ctcggaaggt ggacttcatg 180
ttgatttagc tcaaattatt gaagcctgtg atgtgtgtct gaaggaggat gataaagatg 240
ttgaaagtgt gatgaacagt ggggnatcct actcttgatc cggaanccna c 291

<210> 69
<211> 301
<212> DNA
<213> Homo sapien

<220>

<221> misc_feature
<222> (1)...(301)
<223> n = A,T,C or G

<400> 69
tctatgagca tgccaaggct ctgtgggagg atgaaggagt gcgtgcctgc tacgaacgct 60
ccaacgagta ccagctgatt gactgtgccc agtacttcct ggacaagatc gacgtgatca 120
agcaggctga ctatgtgccg agcgatcagg acctgcttcg ctgccgtgtc ctgacttctg 180
gaatctttga gaccaagttc caggtggacn aagtcaactt ccacatgntt gacgtgggtg 240
gccagcgcga tgaacgccgc aagtggatcc agtgcttcaa cgatgtgact gccatcatct 300
t 301

<210> 70
<211> 201
<212> DNA
<213> Homo sapien

<400> 70
gcggctcttc ctcgggcagc ggaagcggcg cggcggtcgg agaagtggcc taaaacttcg 60
gcgttgggtg aaagaaaatg gcccgaacca agcagactgc tcgtaagtcc accggtggga 120
aagccccccg caaacagctg gccacgaaag ccgccaggaa aagcgctccc tctaccggcg 180
gggtgaagaa gcctcatcgc t 201

<210> 71
<211> 301
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(301)
<223> n = A,T,C or G

<400> 71
gccggggtag tcgccgncgc cgccgccgct gcagccactg caggcaccgc tgccgccgcc 60
tgagtagtgg gcttaggaag gaagaggtca tctcgctcgg agcttcgctc ggaagggtct 120
ttgttccctg cagccctccc acgggaatga caatggataa aagtgagctg gtacanaaag 180
ccaaactcgc tgagcaggct gagcgatatg atgatatggc tgcagccatg aaggcagtca 240
cagaacaggg gcatgaactc ttcaacgaag agagaaatct gctctctggt gcctacaaga 300
a 301

<210> 72
<211> 251
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(251)
<223> n = A,T,C or G

<400> 72
cttggggggt gttgggggag agactgtggg cctggaaata aaacttgtct cctctaccac 60
caccctgtac cctagcctgc acctgtccac atctctgcaa agttcagctt ccttccccag 120
gtctctgtgc actctgtctt ggatgctctg gggagctcat gggtggagga gtctccacca 180
gagggaggct caggggactg gttgggccag ggatgaatat ttgagggata aaaattgtgt 240
aagagccaan g 251

<210> 73
<211> 895
<212> DNA
<213> Homo sapien

<400> 73
```
ttttttttt tttttcccag gccctctttt tatttacagt gataccaaac catccacttg 60
caaattcttt ggtctcccat cagctggaat taagtaggta ctgtgtatct ttgagatcat 120
gtatttgtct ccactttggt ggatacaaga aaggaaggca cgaacagctg aaaaagaagg 180
gtatcacacc gctccagctg gaatccagca ggaacctctg agcatgccac agctgaacac 240
ttaaaagagg aaagaaggac agctgctctt catttatttt gaaagcaaat tcatttgaaa 300
gtgcataaat ggtcatcata agtcaaacgt atcaattaga ccttcaacct aggaaacaaa 360
attttttttt tctatttaat aatacaccac actgaaatta tttgccaatg aatcccaaag 420
atttggtaca aatagtacaa ttcgtatttg ctttcctctt tcctttcttc agacaaacac 480
caaataaaat gcaggtgaaa gagatgaacc acgactagag gctgacttag aaatttatgc 540
tgactcgatc taaaaaaaat tatgttggtt aatgttaatc tatctaaaat agagcatttt 600
gggaatgctt ttcaaagaag gtcaagtaac agtcatacag ctagaaaagt ccctgaaaaa 660
aagaattgtt aagaagtata ataacctttt caaaacccac aatgcagctt agttttcctt 720
tatttatttg tggtcatgaa gactatcccc atttctccat aaaatcctcc ctccatactg 780
ctgcattatg gcacaaaaga ctctaagtgc caccagacag aaggaccaga gtttctgatt 840
ataaacaatg atgctgggta atgtttaaat gagaacattg gatatggatg gtcag 895
```

<210> 74
<211> 351
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(351)
<223> n = A,T,C or G

<400> 74
```
tgtgcncagg ggatgggtgg gcngtggaga ngatgacaga aaggctggaa ggaangggg 60
tgggtttgaa ggccanggcc aaggggncct caggtccgnt tctgnnaagg gacagccttg 120
aggaaggagn catggcaagc catagctagg ccaccaatca gattaagaaa nnctgagaaa 180
nctagctgac catcactgtt ggtgnccagt ttcccaacac aatggaatnc caccacactg 240
gactagngga nccactagtt ctagagcggc cgccaccgcg gtggaacccc aacttttgcc 300
cctttagnga gggttaattg cgcgcttggc ntaatcatgg tcataagctg t 351
```

<210> 75
<211> 251
<212> DNA
<213> Homo sapien

<400> 75
```
tacttgacct tctttgaaaa gcattcccaa aatgctctat tttagataga ttaacattaa 60
ccaacataat ttttttttaga tcgagtcagc ataaatttct aagtcagcct ctagtcgtgg 120
ttcatctctt tcacctgcat tttatttggt gtttgtctga agaaaggaaa gaggaaagca 180
aatacgaatt gtactatttg taccaaatct ttgggattca ttggcaaata atttcagtgt 240
ggtgtattat t 251
```

<210> 76
<211> 251
<212> DNA
<213> Homo sapien

<400> 76
```
tatttaataa tacaccacac tgaaattatt tgccaatgaa tcccaaagat ttggtacaaa 60
tagtacaatt cgtatttgct ttcctctttc ctttcttcag acaaacacca aataaaatgc 120
aggtgaaaga gatgaaccac gactagaggc tgacttagaa atttatgctg actcgatcta 180
aaaaaaatta tgttggttaa tgttaatcta tctaaaatag agcattttgg gaatgctttt 240
caaagaaggt c 251
```

<210> 77
<211> 351

<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(351)
<223> n = A,T,C or G

<400> 77

```
actcaccgtg ctgtgtgctg tgtgcctgct gcctggcagc ctggccctgc cgctgctcag      60
gaggcgggag gcatgagtga gctacagtgg gaacaggctc aggactatct caagagannn     120
tatctctatg actcagaaac aaaaaatgcc aacagtttag aagccaaact caaggagatg     180
caaaaattct ttggcctacc tataactgga atgttaaact cccgcgtcat agaaataatg     240
cagaagccca gatgtggagt gccagatgtt gcagaatact cactatttcc aaatagccca     300
aaatggactt ccaaagtggt cacctacagg atcgtatcat atactcgaga c              351
```

<210> 78
<211> 1574
<212> DNA
<213> Homo sapien

<400> 78

```
gccctggggg cggaggggag gggcccacca cggccttatt tccgcgagcg ccggcactgc      60
ccgctccgag cccgtgtctg tcgggtgccg agccaacttt cctgcgtcca tgcagccccg     120
ccggcaacgg ctgcccgctc cctggtccgg gcccaggggc ccgcgcccca ccgccccgct     180
gctcgcgctg ctgctgttgc tcgccccggt ggcggcgccc gcggggtccg ggaccccga      240
cgaccctggg cagcctcagg atgctggggt cccgcgcagg ctcctgcagc aggcggcgcg     300
cgcggcgctt cacttcttca acttccggtc cggctcgccc agcgcgctgc gagtgctggc     360
cgaggtgcag gagggccgcg cgtggattaa tccaaaagag ggatgtaaag ttcacgtggt     420
cttcagcaca gagcgctaca acccagagtc tttacttcag gaaggtgagg gacgtttggg     480
gaaatgttct gctcgagtgt ttttcaagaa tcagaaaccc agaccaacta tcaatgtaac     540
ttgtacacgg ctcatcgaga aaaagaaaag acaacaagag gattacctgc tttacaagca     600
aatgaagcaa ctgaaaaacc ccttggaaat agtcagcata cctgataatc atggacatat     660
tgatccctct ctgagactca tctgggattt ggctttcctt ggaagctctt acgtgatgtg     720
ggaaatgaca acacaggtgt cacactacta cttggcacag ctcactagtg tgaggcagtg     780
gaaaactaat gatgatacaa ttgattttga ttatactgtt ctacttcatg aattatcaac     840
acaggaaata attccctgtc gcattcactt ggtctggtac cctggcaaac ctcttaaagt     900
gaagtaccac tgtcaagagc tacagacacc agaagaagcc tccggaactg aagaaggatc     960
agctgtagta ccaacagagc ttagtaattt ctaaaaagaa aaaatgatct ttttccgact    1020
tctaaacaag tgactatact agcataaatc attcttctag taaaacagct aaggtataga    1080
cattctaata atttgggaaa acctatgatt acaagtaaaa actcagaaat gcaaagatgt    1140
tggttttttg tttctcagtc tgctttagct tttaactctg gaagcgcatg cacactgaac    1200
tctgctcagt gctaaacagt caccagcagg ttcctcaggg tttcagccct aaaatgtaaa    1260
acctggataa tcagtgtatg ttgcaccaga atcagcattt tttttttaac tgcaaaaaat    1320
gatggtctca tctctgaatt tatatttctc attcttttga acatactata gctaatatat    1380
tttatgttgc taaattgctt ctatctagca tgttaaacaa agataatata ctttcgatga    1440
aagtaaatta taggaaaaaa attaactgtt ttaaaaagaa cttgattatg ttttatgatt    1500
tcaggcaagt attcattttt aacttgctac ctacttttaa ataaatgttt acatttctaa    1560
aaaaaaaaaa aaaa                                                      1574
```

<210> 79
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(401)
<223> n = A,T,C or G

<400> 79

```
catactgtga attgttcttg actccttttc ttgacattca gttttcanaa tttccatctt      60
```

```
tcttctggaa ctaatgtgct gttctcttga ctgcctgctg ggccagcatc cgattgccag    120
ccagaaacgt cacactgccc aagatggcca ggtacttcaa ggtctggaac atgttgagct    180
gagtccagta gacatacatg agtcccagca tagcagcatg tcccaggtga aatataatcg    240
tgctaggagc aaaagtgaag ttggagacat tggcaccaat ccggatccac tagttctaga    300
gcggccgcca ccgcggtgga gctccagctt ttgttccctt tagtgagggt taattgcgcg    360
cttggcgtaa tcatggncat agctgtttcc tgtgtgaaat t                         401

      <210> 80
      <211> 301
      <212> DNA
      <213> Homo sapien

      <400> 80
aaaaatgaaa catctatttt agcagcaaga ggctgtgagg gatggggtag aaaaggcatc     60
ctgagagagt tctagaccga cccaggtcct gtggcacact atacgggtca ggaggggtgg    120
aagacaggcc taagctctag gacggtgaat ctcggggcta tttgtggatt tgttagaaac    180
agacattctt ttggcctttt cctggcactg gtgttgccgg caggtgggca gaagtgagcc    240
accagtcact gttcagtcat tgccaccaca gatcttcagc agaatcttcc ggtaatcccc    300
t                                                                     301

      <210> 81
      <211> 301
      <212> DNA
      <213> Homo sapien

      <220>
      <221> misc_feature
      <222> (1)...(301)
      <223> n = A,T,C or G

      <400> 81
tagccaggtt gctcaagcta attttattct ttcccaacag gatccatttg gaaaatatca     60
agcctttaga atgtggcagc aagagaaagc ggactacgca ggaacgggga gtttgggaga    120
agctctcctg gtgttgactt agggatgaag gctccaggct gctgccagaa atggagtcac    180
cagcagaaga actgntttct ctgataagga tgtcccacca ttttcaagct gttcgttaaa    240
gttacacagg tccttcttgc agcagtaagt accgttagct cattttccct caagcgggtt    300
t                                                                     301

      <210> 82
      <211> 201
      <212> DNA
      <213> Homo sapien

      <220>
      <221> misc_feature
      <222> (1)...(201)
      <223> n = A,T,C or G

      <400> 82
tcaacagaca aaaaaagttt attgaataca aaactcaaag gcatcaacag tcctgggccc     60
aagagatcca tggcaggaag tcaagagttc tgcttcaggg tcggtctggg cagccctgga    120
agaagtcatt gcacatgaca gtgatgagtg ccaggaaaac agcatactcc tggaaagtcc    180
acctgctggn cactgnttca t                                               201

      <210> 83
      <211> 251
      <212> DNA
      <213> Homo sapien

      <220>
      <221> misc_feature
      <222> (1)...(251)
```

<223> n = A,T,C or G

<400> 83
gtaaggagca tactgtgccc atttattata gaatgcagtt aaaaaaaata ttttgaggtt 60
agcctctcca gtttaaaagc acttaacaag aaacacttgg acagcgatgc aatggtctct 120
cccaaaccgg ctccctctta ccaagtaccg taaacagggt ttgagaacgt tcaatcaatt 180
tcttgatatg aacaatcaaa gcatttaatg caaacatatt tgcttctcaa anaataaaac 240
cattttccaa a 251

<210> 84
<211> 301
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(301)
<223> n = A,T,C or G

<400> 84
agtttataat gttttactat gatttagggc tttttttca aagaacaaaa attataagca 60
taaaaactca ggtatcagaa agactcaaaa ggctgttttt cactttgttc agattttgtt 120
tccaggcatt aagtgtgtca tacagttgtt gccactgctg ttttccaaat gtccgatgtg 180
tgctatgact gacaactact tttctctggg tctgatcaat tttgcagtan accattttag 240
ttcttacggc gtcnataaca aatgcttcaa catcatcagc tccaatctga agtcttgctg 300
c 301

<210> 85
<211> 201
<212> DNA
<213> Homo sapien

<400> 85
tatttgtgta tgtaacattt attgacatct acccactgca agtatagatg aataagacac 60
agtcacacca taaaggagtt tatccttaaa aggagtgaaa gacattcaaa aaccaactgc 120
aataaaaaag ggtgacataa ttgctaaatg gagtggagga acagtgctta tcaattcttg 180
attgggccac aatgatatac c 201

<210> 86
<211> 301
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(301)
<223> n = A,T,C or G

<400> 86
tttataaaat attttattta cagtagagct ttacaaaaat agtcttaaat taatacaaat 60
ccctttttgca atataactta tatgactatc ttctcaaaaa cgtgacattc gattataaca 120
cataaactac atttatagtt gttaagtcac cttgtagtat aaatatgttt tcatcttttt 180
tttgtaataa ggtacatacc aataacaatg aacaatggac aacaaatctt attttgntat 240
tcttccaatg taaaattcat ctctggccaa aacaaaatta accaaagaaa agtaaaacaa 300
t 301

<210> 87
<211> 351
<212> DNA
<213> Homo sapien

<220>

<221> misc_feature
<222> (1)...(351)
<223> n = A,T,C or G

<400> 87
aaaaaagatt taagatcata aataggtcat tgttgtcaca acacatttca gaatcttaaa 60
aaaacaaaca ttttggcttt ctaagaaaaa gacttttaaa aaaaatcaat tccctcatca 120
ctgaaaggac ttgtacattt ttaaacttcc agtctcctaa ggcacagtat ttaatcagaa 180
tgccaatatt accaccctgc tgtagcanga ataaagaagc aagggattaa cacttaaaaa 240
aacngccaaa ttcctgaacc aaatcattgg cattttaaaa aagggataaa aaaacnggnt 300
aagggggga gcatttaag taaagaangg ccaagggtgg tatgccngga c 351

<210> 88
<211> 301
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1)...(301)
<223> n = A,T,C or G

<400> 88
gttttaggtc tttaccaatt tgattggttt atcaacaggg catgaggttt aaatatatct 60
ttgaggaaag gtaaagtcaa atttgacttc ataggtcatc ggcgtcctca ctcctgtgca 120
ttttctggtg gaagcacaca gttaattaac tcaagtgtgg cgntagcgat gctttttcat 180
ggngtcattt atccacttgg tgaacttgca cacttgaatg naaactcctg ggtcattggg 240
ntggccgcaa gggaaaggtc cccaagacac caaaccttgc agggtacctn tgcacaccaa 300
c 301

<210> 89
<211> 591
<212> DNA
<213> Homo sapien

<400> 89
tttttttttt ttttttttatt aatcaaatga ttcaaaacaa ccatcattct gtcaatgccc 60
aagcacccag ctggtcctct ccccacatgt cacactctcc tcagcctctc ccccaaccct 120
gctctccctc ctcccctgcc ctagcccagg gacagagtct aggaggagcc tggggcagag 180
ctggaggcag gaagagagca ctggacagac agctatggtt tggattgggg aagagattag 240
gaagtaggtt cttaaagacc ctttttagt accagatatc cagccatatt cccagctcca 300
ttattcaaat catttcccat agcccagctc ctctctgttc tccccctact accaattctt 360
tggctcttac acaattttta tccctcaaat attcatccct ggcccaacca gtcccctgag 420
cctccctctg gtggagactc ctccacccat gagctcccca gagcatccaa gacagagtgc 480
acagagacct ggggaaggaa gctgaacttt gcagagatgt ggacaggtgc aggctagggt 540
acagggtggt ggtagaggag acaagtttta tttccaggcc cacagtctct c 591

<210> 90
<211> 1978
<212> DNA
<213> Homo sapien

<400> 90
tttttttttt tttttttatca aatgaatact ttattagaga cataacacgt ataaaataaa 60
tttcttttca tcatggagtt accagatttt aaaaccaacc aacactttct catttttaca 120
gctaagacat gttaaattct taaatgccat aatttttgtt caactgcttt gtcattcaac 180
tcacaagtct agaatgtgat taagctacaa atctaagtat tcacagatgt gtcttaggct 240
tggtttgtaa caatctagaa gcaatctgtt tacaaaagtg ccaccaaagc attttaaaga 300
aaccaattta atgccaccaa acataagcct gctatacctg ggaaacaaaa aatctcacac 360
ctaaattcta gcagagtaaa cgattccaac tagaatgtac tgtatatcca tatggcacat 420
ttatgacttt gtaatatgta attcataata caggtttagg tgtgtggtat ggagctagga 480
aaaccaaagt agtaggatat tatagaaaag atctgatgtt aagtataaag tcatatgcct 540

```
gatttcctca aacctttttgt ttttcctcat gtcttctgtc tttatatttt tatcacaaac     600
caagatctaa cagggttctt tctagaggat tattagataa gtaacacttg atcattaagc     660
acggatcatg ccactcattc atggttgttc tatgttccat gaactctaat agcccaactt     720
atacatggca ctccaagggg atgcttcagc cagaaagtaa agggctgaaa aagtagaaca     780
atacaaaagc cctcgtgtgg tgggaactgt ggcctcactc ttacttgtcc ttccattcaa     840
aacagtttgg cacctttcca tgacgaggat ctctacaggt aggttaaaat acttttctgt     900
gctattcagc cagaaatagt ttttgtgctg gatatgattt taaaacagat tttgtctgtc     960
accagtgcaa aaacattaca gatgtctggg ctaatacaaa aacacataag aatctacaac    1020
tttatattta atactctatt caaatttaac tcaaagtaat gcaaaataat tagaagtaaa    1080
aacttaattc ttctgagagc tctatttgga aaagcttcac atatccacac acaaatatgg    1140
gtatattcat gcacagggca aacaactgta ttctgaagca taaataaact caaagtaaga    1200
catcagtagc tagataccag ttccagtatt ggttaatggt ctctggggat cccattttaa    1260
gcactctcag atgaggatct tgctcagttg ttagactatc attagtttga ttaagcaact    1320
gaagtttact tcataaatta ctttttccta tatccaggac tctgcctgag aaatttttata    1380
cattcctcca aaggtaagta ttctccaaag gtaagtattt gactattaac acaaaggcaa    1440
tgtgattatt gcataatgac actaaatatt atgtggcttt tctgttaggt ttataagttt    1500
tcaatgatca gttcaagaaa atgcagatca tatataacta aggttttaca ccagtggttg    1560
acaaactatg gcccacaggc taaacccagc ctccccttgt ttttataaat aagttttatt    1620
agacataacc acactcattc atttctgtat tgtgtatagc tgctttcacg ctatactagc    1680
agaactgaat agttgtgaca gagactgtat ggaccgtgaa gcataaatat ttaccatctg    1740
gcccattcta aaaaaagtgt gccaattcct ggtttacact aaaatataga gtttagtggg    1800
aagcctattt gaaatgtgtt tttttaggg gctgtaatta ccaattaaaa ttaaggttca    1860
ggtgactcag caaccaaaca aaagggatac taatttttta tgaacaatat atttgtattt    1920
tatggacata aaggaaact ttcagaaaga aaaggaggaa aataaagggg gaaaggga      1978
```

<210> 91
<211> 895
<212> DNA
<213> Homo sapien

<400> 91

```
tttttttttt tttttcttg tttaaaaaaa ttgttttcat tttaatgatc tgagttagta       60
acaaacaaat gtacaaaatt gtctttcaca tttccataca ttgtgttatg gaccaaatga     120
aaacgctgga ctacaaatgc aggtttcttt atatccttaa cttcaattat tgtcacttat     180
aaataaaggt gatttgctaa cacatgcatt tgtgaacaca gatgccaaaa attatacatg     240
taagttaatg cacaaccaag agtatacact gttcatttgt gcagttatgc gtcaaatgcg     300
actgacacag aagcagttat cctgggatat ttcactctat atgaaaagca tcttggagaa     360
atagattgaa atacagttta aaacaaaaat tgtattctac aaatacaata aaatttgcaa     420
cttgcacatc tgaagcaaca tttgagaaag ctgcttcaat aaccctgctg ttatattggt     480
tttataggta tatctccaaa gtcatgggtt gggatatagc tgctttaaag aaaataaata     540
tgtatattaa aaggaaaatc acactttaaa aatgtgagga aagctttgaa aacagtctta     600
atgcatgagt ccatctacat attttcaagt tttggaaaca gaaagaagtt tagaattttc     660
aaagtaatct gaaaactttc taagccattt taaaataaga tttttttccc catctttcca     720
atgtttccta tttgatagtg taatacagaa atgggcagtt tctagtgtca acttaactgt     780
gctaattcat aagtcattat acatttatga cttaagagtt caaataagtg gaaattgggt     840
tataatgaaa atgacaaggg ggccccttca gcagccactc atctgaacta gtaat          895
```

<210> 92
<211> 1692
<212> DNA
<213> Homo sapien

<400> 92

```
tttttttttt tttttaactt ttagcagtgt ttatttttgt taaaagaaac caattgaatt      60
gaaggtcaag acaccttctg attgcacaga ttaaacaaga aagtattact tatttcaact     120
ttacaaagca tcttattgat ttaaaaagat ccatactatt gataaagttc accatgaaca     180
tatatgtaat aaggagacta aaatattcat tttacatatc tacaacatgt atttcatatt     240
tctaatcaac cacaaatcat ataggaaaat atttaggtcc atgaaaaagt ttcaaaacat     300
taaaaaatta aagttttgaa acaaatcaca tgtgaaagct cattaaataa taacattgac     360
aaataaatag ttaatcagct ttacttatta gctgctgcca tgcatttctg gcattccatt     420
ccaagcgagg gtcagcatgc agggtataat ttcatactat gcgaccgtaa agagctacag     480
ggcttatttt tgaagtgaaa tgtcacaggg tctttcattc tctttcaaag gaagatcact     540
```

catggctgct aaactgttcc catgaagagt accaaaaaag cacctttctg aaatgttact 600
gtgaagattc atgacaacat attttttta acctgttttg aaggagtttt gtttaggaga 660
ggggatgggc cagtagatgg agggtatctg agaagccctt ttctgtttta aaatataatg 720
attcactgat gtttatagta tcaacagtct tttaagaaca atgaggaatt aaaactacag 780
gatacgtgga atttaaatgc aaattgcatt catggatata cctacatctt gaaaaacttg 840
aaaaggaaaa actattccca aagaaggtcc tgatacttaa gacagcttgc tgggtttgat 900
caaagcagaa agcatatact ttcaagtgag aaaacagcag tggcaggctt gagtcttcca 960
agcaatcaaa tctgtaaagc agatggttac tagtaagtct agttatggga gtctgagttc 1020
taactcatgc tgtgcttgct ggatttgctg gctcttttcc gctctctgtg atgctggact 1080
ggcttgtcag gtgacatgct ctcaaagttg tgactggact cgttgtgctg ccgggtgtac 1140
ctcttgcact tgcaggcagt gactactgtg attttgtagg tgcgtgtgct gccatcttgg 1200
cactgcagct ggattctctg ggtacgggtt ttgtcattga cacaccgcca ctcctgggag 1260
ctcctcctgc tccagtactt tgttccatag cctcctccaa tccagttagg gagcactggc 1320
aggggcaagc actcgccagc acacaccagc tccttcagag ggctgatgct ggtgcactgg 1380
ccatcagaga tgtatttggt ggaacgcagt tcccggcaac ccacttgaac ccgagtgttc 1440
cgatccagtc cagtgttact gaaatgcctg cctccatttc tggcttgatt caacgtgctg 1500
ttgctgctgg ggtgtgctgg aacaggttta accacatgtg aataaaggat ttctgtggca 1560
tcatttttaa aagccaaaca gcttttcatt aggatgcatg caaggggaag gagatagaaa 1620
tgaatggcag gaggaagcat ggtgagtaga ggatttgctt gactgaagag ctggttaatt 1680
cttttgcctc tg 1692

<210> 93
<211> 251
<212> DNA
<213> Homo sapien

<400> 93
cccaccctac ccaaatatta gacaccaaca cagaaaagct agcaatggat tcccttctac 60
tttgttaaat aaataagtta aatatttaaa tgcctgtgtc tctgtgatgg caacagaagg 120
accaacaggc cacatcctga taaaaggtaa gaggggggtg gatcagcaaa aagacagtgc 180
tgtgggctga ggggacctgg ttcttgtgtg ttgcccctca agactcttcc cctacaaata 240
actttcatat g 251

<210> 94
<211> 735
<212> DNA
<213> Homo sapien

<400> 94
tttttttttt tttttccact tctcagttta tttctgggac taaatttggg tcagagctgc 60
agagaaggga tgggccctga gcttgaggat gaaagtgccc cagggagatt gagacgcaac 120
ccccgccctg gacagttttg gaaattgttc ccagggttca actagagaga cacggtcagc 180
ccaatgtggg ggaagcagac cctgagtcca ggagacatgg ggtcaggggc tggagagatg 240
aacattctca acatctctgg gaaggaatga gggtctgaaa ggagtgtcag ggctgtccct 300
gcagcaggtg gggatgccgg tgtgctgagt cctgggatga ctcaggagtt ggcctggatg 360
gtttcctgga tccacttggt gaacttgcag aggttcgtgt agacacccgg tctgttgggc 420
cgggcacaag ggtaatctcc ccaggacacg agtccctgca gggagccatt gcagaccaca 480
ggcccccag aatcaccctg gcaggagtct ctacctgctt tgtcaccggc gcagaacatg 540
gtgtcatcta tctgtctcgg gtaagcatcc tcgcaccttt tctgacttag cacgctgata 600
ttcaagcact ggaggacctt agggaagtgc acttgggggc tcttggttgt cccccagcca 660
gacaccaagc actttgtccc agcagaggga caatgagagg agacgttgat gggtctgaca 720
tctttagtgg gacga 735

<210> 95
<211> 578
<212> DNA
<213> Homo sapien

<400> 95
cttgccttct cttaggcttt gaagcatttt tgtctgtgct ccctgatctt caggtcacca 60
ccatgaagtt cttagcagtc ctggtactct tgggagtttc catctttctg gtctctgccc 120
agaatccgac aacagctgct ccagctgaca cgtatccagc tactggtcct gctgatgatg 180

```
aagcccctga tgctgaaacc actgctgctg caaccactgc gaccactgct gctcctacca     240
ctgcaaccac cgctgcttct accactgctc gtaaagacat tccagtttta cccaaatggg     300
ttggggatct cccgaatggt agagtgtgtc cctgagatgg aatcagcttg agtcttctgc     360
aattggtcac aactattcat gcttcctgtg atttcatcca actacttacc ttgcctacga     420
tatcccctt  atctctaatc agtttatttt ctttcaaata aaaaataact atgagcaaca     480
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa     540
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa                             578
```

<210> 96
<211> 594
<212> DNA
<213> Homo sapien

<400> 96

```
atggcaaaga atggacttgt aatttgcatc ctggtgatca ccttactcct ggaccagacc      60
accagccaca catccagatt aaaagccagg aagcacagca aacgtcgagt gagagacaag     120
gatggagatc tgaagactca aattgaaaag ctctggacag aagtcaatgc cttgaaggaa     180
attcaagccc tgcagacagt ctgtctccga ggcactaaag ttcacaagaa atgctacctt     240
gcttcagaag gtttgaagca tttccatgag gccaatgaag actgcatttc caaaggagga     300
atcctggtta tccccaggaa ctccgacgaa atcaacgccc tccaagacta tggtaaaagg     360
agcctgccag gtgtcaatga cttttggctg ggcatcaatg acatggtcac ggaaggcaag     420
tttgttgacg tcaacggaat cgctatctcc ttcctcaact gggaccgtgc acagcctaac     480
ggtggcaagc gagaaaactg tgtcctgttc tcccaatcag ctcagggcaa gtggagtgat     540
gaggcctgtc gcagcagcaa gagatacata tgcgagttca ccatccctca atag           594
```

<210> 97
<211> 3101
<212> DNA
<213> Homo sapien

<400> 97

```
tgttggggcc tcagcctccc aagtagctgg gactacaggt gcctgccacc acgcccagct      60
aattttttgt atattttta  gtagagacgg ggtttcaccg tggtctcaat ctcctgacct     120
cgtgatctgc cagccttggc ctcccaaagt gtattctctt tttattatta ttattatttt     180
tgagatggag tctgtctctg tcgcccaggc tggagtgcag tggtgcgatc tctgctcact     240
gcaagctccg cctcctgggt tcatgccatt ctcctgcctc agcctcccga gtagctggga     300
ctacaggccc ctgccaccac acccggctaa ttttttgtat ttttagtaga gacaggggttt    360
caccatgtta gccagggtgg tctctatctt ctgacctcgt gatccgcctg cctcagtctc     420
tcaaagtgct gggattacag gcgtgagcca ccgcgaccag ccaactattg ctgtttattt     480
ttaaatatat tttaaagaaa caattagatt tgtttcttt  ctcattcttt tacttctact     540
cttcatgtat gtataattat atttgtgttt tctattacct tttctccttt tactgtattg     600
gactataata attgtgctca ctaatttctg ttcactaata ttatcagctt agataatact     660
ttaattttta acttatatat tgagtattaa attgatcagt tttatttgta attatctatc     720
ttccgcttgg ctgaatataa cttcttaagc ttataacttc ttgttctttc catgttattt     780
ttttcttttt tttaatgtat tgaatttctt ctgacactca ttctagtaac ttttttctcg     840
gtgtgcaacg taagttataa tttgtttctc agatttgaga tctgccataa gtttgaggct     900
ttattttttt tttttatttg ctttatggca agtcggacaa cctgcatgga tttggcatca     960
atgtagtcac ccatatctaa gagcagcact tgcttcttag catgatgagt tgtttctgga    1020
ttgtttcttt attttactta tattcctggt agattcttat atttcccctt caactctatt    1080
cagcatttta ggaattctta ggactttctg agaattttag ctttctgtat taaatgtttt    1140
taatgagtat tgcattttct caaaaagcac aaatatcaat agtgtacaca tgaggaaaac    1200
tatatatata ttctgttgca gatgacagca tctcataaca aaatcctagt tacttcattt    1260
aaaagacagc tctcctccaa tatactatga ggtaacaaaa atttgtagtg tgtaattttt    1320
ttaatattag aaaactcatc ttacattgtg cacaaatttc tgaagtgata atacttcact    1380
gtttttctat agaagtaact taatattggc aaaattactt atttgaattt aggttttggc    1440
tttcatcata tacttcctca ttaacatttc cctcaatcca taaatgcaat ctcagtttga    1500
atcttccatt taacccagaa gttaattttt aaaaccttaa taaaatttga atgtagctag    1560
atattatttg ttggttacat attagtcaat aatttatatt acttacaatg atcagaaaat    1620
atgatctgaa tttctgctgt cataaattca ataacgtatt ttaggcctaa acctttccat    1680
ttcaaatcct tgggtctggt aattgaaaat aatcattatc ttttgttttc tggccaaaaa    1740
tgctgcccat ttatttctat ccctaattag tcaaactttc taataaatgt atttaacgtt    1800
aatgatgttt atttgcttgt tgtatactaa aaccattagt ttctataatt taaatgtcac    1860
```

```
ctaatatgag tgaaaatgtg tcagaggctg gggaagaatg tggatggaga aagggaaggt   1920
gttgatcaaa aagtacccaa gtttcagtta cacaggaggc atgagattga tctagtgcaa   1980
aaaatgatga gtataataaa taataatgca ctgtatattt tgaaattgct aaaagtagat   2040
ttaaaattga tttacataat attttacata tttataaagc acatgcaata tgttgttaca   2100
tgtatagaat gtgcaacgat caagtcaggg tatctgtggt atccaccact ttgagcattt   2160
atcgattcta tatgtcagga acatttcaag ttatctgttc tagcaaggaa atataaaata   2220
cattatagtt aactatggcc tatctacagt gcaactaaac actagatttt attcctttcc   2280
aactgtgggt ttgtattcat ttaccaccct cttttcattc cctttctcac ccacacactg   2340
tgccgggcct caggcatata ctattctact gtctgtctct gtaaggatta tcattttagc   2400
ttccacatat gagagaatgc atgcaaagtt tttctttcca tgtctggctt atttcactta   2460
acaaaatgac ctccgcttcc atccatgtta tttatattac ccaatagtgt tcataaatat   2520
atatacacac atatatacca cattgcattt gtccaattat tcattgacgg aaactggtta   2580
atgttatatc gttgctattg tgaatagtgc tgcaataaac acgcaagtgg ggatataatt   2640
tgaagagttt ttttgttgat gttccataca aattttaaga ttgttttgtc tatgtttgtg   2700
aaaatggcgt tagtattttc atagagattg cattgaatct gtagattgct ttgggtaagt   2760
atggttattt tgatggtatt aatttttca ttccatgaag atgagatgtc tttccatttg   2820
tttgtgtcct ctacattttc tttcatcaaa gttttgttgt atttttgaag tagatgtatt   2880
tcaccttata gatcaagtgt attccctaaa tattttattt ttgtagctat tgtagatgaa   2940
attgccttct cgatttcttt ttcacttaat tcattattag tgtatggaaa tgttatggat   3000
ttttatttgt tggtttttaa tcaaaaactg tattaaactt agagtttttt gtggagtttt   3060
taagtttttc tagatataag atcatgacat ctaccaaaaa a                       3101
```

```
<210> 98
<211> 90
<212> PRT
<213> Homo sapien

<400> 98
Met Lys Phe Leu Ala Val Leu Val Leu Leu Gly Val Ser Ile Phe Leu
 1              5                  10                  15
Val Ser Ala Gln Asn Pro Thr Thr Ala Ala Pro Ala Asp Thr Tyr Pro
            20                  25                  30
Ala Thr Gly Pro Ala Asp Asp Glu Ala Pro Asp Ala Glu Thr Thr Ala
        35                  40                  45
Ala Ala Thr Thr Ala Thr Thr Ala Ala Pro Thr Thr Ala Thr Thr Ala
    50                  55                  60
Ala Ser Thr Thr Ala Arg Lys Asp Ile Pro Val Leu Pro Lys Trp Val
65                  70                  75                  80
Gly Asp Leu Pro Asn Gly Arg Val Cys Pro
                85                  90

<210> 99
<211> 197
<212> PRT
<213> Homo sapien

<400> 99
Met Ala Lys Asn Gly Leu Val Ile Cys Ile Leu Val Ile Thr Leu Leu
 1              5                  10                  15
Leu Asp Gln Thr Thr Ser His Thr Ser Arg Leu Lys Ala Arg Lys His
            20                  25                  30
Ser Lys Arg Arg Val Arg Asp Lys Asp Gly Asp Leu Lys Thr Gln Ile
        35                  40                  45
Glu Lys Leu Trp Thr Glu Val Asn Ala Leu Lys Glu Ile Gln Ala Leu
    50                  55                  60
Gln Thr Val Cys Leu Arg Gly Thr Lys Val His Lys Lys Cys Tyr Leu
65                  70                  75                  80
Ala Ser Glu Gly Leu Lys His Phe His Glu Ala Asn Glu Asp Cys Ile
                85                  90                  95
Ser Lys Gly Gly Ile Leu Val Ile Pro Arg Asn Ser Asp Glu Ile Asn
            100                 105                 110
Ala Leu Gln Asp Tyr Gly Lys Arg Ser Leu Pro Gly Val Asn Asp Phe
```

```
        115                 120                 125
Trp Leu Gly Ile Asn Asp Met Val Thr Glu Gly Lys Phe Val Asp Val
    130                 135                 140
Asn Gly Ile Ala Ile Ser Phe Leu Asn Trp Asp Arg Ala Gln Pro Asn
145                 150                 155                 160
Gly Gly Lys Arg Glu Asn Cys Val Leu Phe Ser Gln Ser Ala Gln Gly
                165                 170                 175
Lys Trp Ser Asp Glu Ala Cys Arg Ser Ser Lys Arg Tyr Ile Cys Glu
            180                 185                 190
Phe Thr Ile Pro Gln
        195

      <210> 100
      <211> 3410
      <212> DNA
      <213> Homo sapien

      <400> 100
gggaaccagc ctgcacgcgc tggctccggg tgacagccgc gcgcctcggc caggatctga       60
gtgatgagac gtgtccccac tgaggtgccc cacagcagca ggtgttgagc atgggctgag      120
aagctggacc ggcaccaaag ggctggcaga aatgggcgcc tggctgattc ctaggcagtt      180
ggcggcagca aggaggagag gccgcagctt ctggagcaga gccgagacga agcagttctg      240
gagtgcctga acggcccccct gagccctacc cgcctggccc actatggtcc agaggctgtg      300
ggtgagccgc ctgctgcggc accggaaagc ccagctcttg ctggtcaacc tgctaacctt      360
tggcctggag gtgtgtttgg ccgcaggcat cacctatgtg ccgcctctgc tgctggaagt      420
gggggtagag gagaagttca tgaccatggt gctgggcatt ggtccagtgc tgggcctggt      480
ctgtgtcccg ctcctaggct cagccagtga ccactggcgt ggacgctatg gccgccgccg      540
gcccttcatc tgggcactgt ccttgggcat cctgctgagc ctctttctca tcccaagggc      600
cggctggcta gcagggctgc tgtgcccgga tcccaggccc ctggagctgg cactgctcat      660
cctgggcgtg gggctgctgg acttctgtgg ccaggtgtgc ttcactccac tggaggccct      720
gctctctgac ctcttccggg acccggacca ctgtcgccag gcctactctg tctatgcctt      780
catgatcagt cttgggggct gcctgggcta cctcctgcct gccattgact gggacaccag      840
tgccctggcc ccctacctgg gcacccagga ggagtgcctc tttggcctgc tcaccctcat      900
cttcctcacc tgcgtagcag ccacactgct ggtggctgag gaggcagcgc tgggccccac      960
cgagccagca gaagggctgt cggcccctc cttgtcgccc cactgctgtc catgccgggc     1020
ccgcttggct ttccggaacc tgggcgccct gcttccccgg ctgcaccagc tgtgctgccg     1080
catgccccgc accctgcgcc ggctcttcgt ggctgagctg tgcagctgga tggcactcat     1140
gaccttcacg ctgttttaca cggatttcgt gggcgagggg ctgtaccagg gcgtgcccag     1200
agctgagccg ggcaccgagg cccggagaca ctatgatgaa ggcgttcgga tgggcagcct     1260
ggggctgttc ctgcagtgcg ccatctccct ggtcttctct ctggtcatgg accggctggt     1320
gcagcgattc ggcactcgag cagtctattt ggccagtgtg gcagctttcc ctgtggctgc     1380
cggtgccaca tgcctgtccc acagtgtggc cgtggtgaca gcttcagccg ccctcaccgg     1440
gttcaccttc tcagccctgc agatcctgcc ctacacactg gcctccctct accaccggga     1500
gaagcaggtg ttcctgccca aataccgagg ggacactgga ggtgctagca gtgaggacag     1560
cctgatgacc agcttcctgc caggccctaa gcctggagct cccttcccta atggacacgt     1620
gggtgctgga ggcagtggcc tgctcccacc tccacccgcg ctctgcgggg cctctgcctg     1680
tgatgtctcc gtacgtgtgg tggtgggtga gcccaccgag gccagggtgg ttccgggccg     1740
gggcatctgc ctggacctcg ccatcctgga tagtgccttc ctgctgtccc aggtggcccc     1800
atccctgttt atgggctcca ttgtccagct cagccagtct gtcactgcct atatggtgtc     1860
tgccgcaggc ctgggtctgg tcgccattta ctttgctaca caggtagtat ttgacaagag     1920
cgacttggcc aaatactcag cgtagaaaac ttccagcaca ttggggtgga gggcctgcct     1980
cactgggtcc cagctccccg ctcctgttag ccccatgggg ctgccgggct ggccgccagt     2040
ttctgttgct gccaaagtaa tgtggctctc tgctgccacc ctgtgctgct gaggtgcgta     2100
gctgcacagc tgggggctgg ggcgtccctc tcctctcc ccagtctcta gggctgcctg     2160
actggaggcc ttccaagggg gtttcagtct ggacttatac agggaggcca gaagggctcc     2220
atgcactgga atgcggggac tctgcaggtg gattacccag gctcagggtt aacagctagc     2280
ctcctagttg agacacacct agagaagggt ttttgggagc tgaataaact cagtcacctg     2340
gtttcccatc tctaagcccc ttaacctgca gcttcgttta atgtagctct tgcatgggag     2400
tttctaggat gaaacactcc tccatgggat ttgaacatat gacttatttg taggggaaga     2460
gtcctgaggg gcaacacaca agaaccaggt cccctcagcc cacagcactg tctttttgct     2520
gatccacccc cctcttacct tttatcagga tgtggcctgt tggtccttct gttgccatca     2580
cagagacaca ggcatttaaa tatttaactt atttatttaa caaagtagaa gggaatccat     2640
```

```
tgctagcttt tctgtgttgg tgtctaatat ttgggtaggg tgggggatcc ccaacaatca    2700
ggtcccctga gatagctggt cattgggctg atcattgcca gaatcttctt ctcctggggt    2760
ctggcccccc aaaatgccta acccaggacc ttggaaattc tactcatccc aaatgataat    2820
tccaaatgct gttacccaag gttagggtgt tgaaggaagg tagagggtgg ggcttcaggt    2880
ctcaacggct tccctaacca cccctcttct cttggcccag cctggttccc cccacttcca    2940
ctcccctcta ctctctctag gactgggctg atgaaggcac tgcccaaaat ttcccctacc    3000
cccaactttc ccctaccccc aactttcccc accagctcca caaccctgtt tggagctact    3060
gcaggaccag aagcacaaag tgcggtttcc caagcctttg tccatctcag cccccagagt    3120
atatctgtgc ttggggaatc tcacacagaa actcaggagc accccctgcc tgagctaagg    3180
gaggtcttat ctctcagggg gggtttaagt gccgtttgca ataatgtcgt cttatttatt    3240
tagcggggtg aatattttat actgtaagtg agcaatcaga gtataatgtt tatggtgaca    3300
aaattaaagg ctttcttata tgtttaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa    3360
aaaaaaaara aaaaaaaaaa aaaaaaaaaa aaaaaaataa aaaaaaaaaa               3410
```

```
<210> 101
<211> 553
<212> PRT
<213> Homo sapien

<400> 101
Met Val Gln Arg Leu Trp Val Ser Arg Leu Leu Arg His Arg Lys Ala
 1               5                  10                  15
Gln Leu Leu Leu Val Asn Leu Leu Thr Phe Gly Leu Glu Val Cys Leu
            20                  25                  30
Ala Ala Gly Ile Thr Tyr Val Pro Pro Leu Leu Leu Glu Val Gly Val
        35                  40                  45
Glu Glu Lys Phe Met Thr Met Val Leu Gly Ile Gly Pro Val Leu Gly
    50                  55                  60
Leu Val Cys Val Pro Leu Leu Gly Ser Ala Ser Asp His Trp Arg Gly
65                  70                  75                  80
Arg Tyr Gly Arg Arg Arg Pro Phe Ile Trp Ala Leu Ser Leu Gly Ile
                85                  90                  95
Leu Leu Ser Leu Phe Leu Ile Pro Arg Ala Gly Trp Leu Ala Gly Leu
            100                 105                 110
Leu Cys Pro Asp Pro Arg Pro Leu Glu Leu Ala Leu Leu Ile Leu Gly
        115                 120                 125
Val Gly Leu Leu Asp Phe Cys Gly Gln Val Cys Phe Thr Pro Leu Glu
    130                 135                 140
Ala Leu Leu Ser Asp Leu Phe Arg Asp Pro Asp His Cys Arg Gln Ala
145                 150                 155                 160
Tyr Ser Val Tyr Ala Phe Met Ile Ser Leu Gly Gly Cys Leu Gly Tyr
                165                 170                 175
Leu Leu Pro Ala Ile Asp Trp Asp Thr Ser Ala Leu Ala Pro Tyr Leu
            180                 185                 190
Gly Thr Gln Glu Glu Cys Leu Phe Gly Leu Leu Thr Leu Ile Phe Leu
        195                 200                 205
Thr Cys Val Ala Ala Thr Leu Leu Val Ala Glu Glu Ala Ala Leu Gly
    210                 215                 220
Pro Thr Glu Pro Ala Glu Gly Leu Ser Ala Pro Ser Leu Ser Pro His
225                 230                 235                 240
Cys Cys Pro Cys Arg Ala Arg Leu Ala Phe Arg Asn Leu Gly Ala Leu
                245                 250                 255
Leu Pro Arg Leu His Gln Leu Cys Cys Arg Met Pro Arg Thr Leu Arg
            260                 265                 270
Arg Leu Phe Val Ala Glu Leu Cys Ser Trp Met Ala Leu Met Thr Phe
        275                 280                 285
Thr Leu Phe Tyr Thr Asp Phe Val Gly Glu Gly Leu Tyr Gln Gly Val
    290                 295                 300
Pro Arg Ala Glu Pro Gly Thr Glu Ala Arg Arg His Tyr Asp Glu Gly
305                 310                 315                 320
Val Arg Met Gly Ser Leu Gly Leu Phe Leu Gln Cys Ala Ile Ser Leu
                325                 330                 335
```

```
Val Phe Ser Leu Val Met Asp Arg Leu Val Gln Arg Phe Gly Thr Arg
            340                 345                 350
Ala Val Tyr Leu Ala Ser Val Ala Ala Phe Pro Val Ala Ala Gly Ala
        355                 360                 365
Thr Cys Leu Ser His Ser Val Ala Val Val Thr Ala Ser Ala Ala Leu
    370                 375                 380
Thr Gly Phe Thr Phe Ser Ala Leu Gln Ile Leu Pro Tyr Thr Leu Ala
385                 390                 395                 400
Ser Leu Tyr His Arg Glu Lys Gln Val Phe Leu Pro Lys Tyr Arg Gly
                405                 410                 415
Asp Thr Gly Gly Ala Ser Ser Glu Asp Ser Leu Met Thr Ser Phe Leu
            420                 425                 430
Pro Gly Pro Lys Pro Gly Ala Pro Phe Pro Asn Gly His Val Gly Ala
        435                 440                 445
Gly Gly Ser Gly Leu Leu Pro Pro Pro Pro Ala Leu Cys Gly Ala Ser
    450                 455                 460
Ala Cys Asp Val Ser Val Arg Val Val Val Gly Glu Pro Thr Glu Ala
465                 470                 475                 480
Arg Val Val Pro Gly Arg Gly Ile Cys Leu Asp Leu Ala Ile Leu Asp
                485                 490                 495
Ser Ala Phe Leu Leu Ser Gln Val Ala Pro Ser Leu Phe Met Gly Ser
            500                 505                 510
Ile Val Gln Leu Ser Gln Ser Val Thr Ala Tyr Met Val Ser Ala Ala
        515                 520                 525
Gly Leu Gly Leu Val Ala Ile Tyr Phe Ala Thr Gln Val Val Phe Asp
    530                 535                 540
Lys Ser Asp Leu Ala Lys Tyr Ser Ala
545                 550

      <210> 102
      <211> 940
      <212> DNA
      <213> Human

      <400> 102
tttactgctt ggcaaagtac cctgagcatc agcagagatg ccgagatgaa atcagggaac      60
tcctagggga tgggtcttct attacctggg aacacctgag ccagatgcct tacaccacga     120
tgtgcatcaa ggaatgcctc cgcctctacg caccggtagt aaacatatcc cggttactcg     180
acaaacccat caccttcca gatggacgct ccttacctgc aggaataact gtgtttatca      240
atatttgggc tcttcaccac aacccctatt tctgggaaga ccctcaggtc tttaacccct     300
tgagattctc cagggaaaat tctgaaaaaa tacatcccta tgccttcata ccattctcag     360
ctggattaag gaactgcatt gggcagcatt ttgccataat tgagtgtaaa gtggcagtgg     420
cattaactct gctccgcttc aagctggctc cagaccactc aaggcctccc cagcctgttc     480
gtcaagttgt cctcaagtcc aagaatggaa tccatgtgtt tgcaaaaaaa gtttgctaat     540
tttaagtcct ttcgtataag aattaatgag acaattttcc taccaaagga agaacaaaag     600
gataaatata atacaaaata tatgtatatg gttgtttgac aaattatata acttaggata     660
cttctgactg gttttgacat ccattaacag taattttaat ttctttgctg tatctggtga     720
aacccacaaa aacmcctgaa aaaactcaag ctgacttcca ctgcgaaggg aaattattgg     780
tttgtgtaac tagtggtaga gtggctttca agcatagttt gatcaaaact ccactcagta     840
tctgcattac ttttatytyt gcaaatatct gcatgatagc tttattytca gttatctttc     900
cccataataa aaaatatctg ccaaaaaaaa aaaaaaaaaa                           940

      <210> 103
      <211> 529
      <212> DNA
      <213> Human

      <400> 103
tttttttttt tttttactga tagatggaat ttattaagct tttcacatgt gatagcacat      60
agttttaatt gcatccaaag tactaacaaa aactctagca atcaaraatg gcagcatgtt     120
attttataac aatcaacacc tgtggctttt aaaatttggt tttcataara taatttatac     180
tgaagtaaat ctagccatgc ttttaaaaaa tgctttaggt cactccaagc ttggcagtta     240
```

```
acatttggca taaacaataa taaaacaatc acaatttaat aaataacaaa tacaacattg     300
taggccataa tcatatacag tataaggaaa aggkggtagt gttgagtaag cagttattag     360
aatagaatac cttggcctct atgcaaatat gtctaracac tttgattcac tcagccctga     420
cattcagttt tcaaagtagg agacaggttc tacagtatca ttttacagtt tccaacacat     480
tgaaaacaag tagaaaatga tgagttgatt tttattaatg cattacatc                 529

      <210> 104
      <211> 469
      <212> DNA
      <213> Human

      <400> 104
cccaacacaa tggataaaaa cacttatagt aaatggggac attcactata atgatctaag      60
aagctacaga ttgtcatagt tgttttcctg ctttacaaaa ttgctccaga tctggaatgc     120
cagtttgacc tttgtcttct ataatatttc cttttttcc cctctttgaa tctctgtata      180
tttgattctt aactaaaatt gttctcttaa atattctgaa tcctggtaat taaaagtttg     240
ggtgtatttt ctttacctcc aaggaaagaa ctactagcta caaaaaatat tttggaataa     300
gcattgtttt ggtataaggt acatattttg gttgaagaca ccagactgaa gtaaacagct     360
gtgcatccaa tttattatag ttttgtaagt aacaatatgt aatcaaactt ctaggtgact     420
tgagagtgga acctcctata tcattattta gcaccgtttg tgacagtaa                 469

      <210> 105
      <211> 744
      <212> DNA
      <213> Human

      <400> 105
ggcctgggac aggattgagg tatgttgcag cctccagggc ctggggtctc ctgcatgaag      60
aatacccctc cccatttgac tgtgaacttt ttggcctgga ttctggagaa cagatttcca     120
ggattgtcag ccagaaggca gacagatgca ggcacctacc aagacctgac ctcaggaagt     180
ggccctgccc tacagcccag ttgctcagcc agggctgaag gccatggggc cccagcaccc     240
ttgcttcagt gccagcccct ggaaggaacc tcacaacagg gatacagcaa ggacactcca     300
gttcccccag tcctgccatg gtgctaccct gagggacagg gatggagaca gggcagccag     360
gtttgccagg acctgcatag cgggcccaag actgcccttc ctcttaagtc atgccaaagc     420
ctccctgccc agtctgagac agtcgctggc aggtgaccac gacctgcgtg gccctcccgg     480
cagttgtcat ggtggttgta ccccacccca tccccctgag gagacatggg ctcagtccca     540
tgcctggtgc ccacagccac aaagatggcc atgggtctct agcctgatat tcgtggcctg     600
gcaggggtca gcacccctga gggcatccaa gccatggtca gaggaaagtg ttggcaggct     660
cggcacagcc aaagaagtca ggacccacga gacgggggaa gccttccaga gccttcacct     720
tcacagggtc aaacttccag taga                                            744

      <210> 106
      <211> 401
      <212> DNA
      <213> Human

      <400> 106
acattgttag gtgctgacct agacagagat gaactgaggt ccttgttttg ttttgttcat      60
aatacaaagg tgctaattaa tagtatttca gatacttgaa gaatgttgat ggtgctagaa     120
gaatttgaga agaaatactc ctgtattgag ttgtatcgtg tggtgtattt tttaaaaaat     180
ttgatttagc attcatattt tccatcttat tcccaattaa aagtatgcag attatttgcc     240
caaatcttct tcagattcag catttgttct ttgccagtct cattttcatc ttcttccatg     300
gttccacaga agctttgttt cttgggcaag cagaaaaatt aaattgtacc tattttgtat     360
atgtgagatg tttaaataaa ttgtgaaaaa aatgaaataa a                         401

      <210> 107
      <211> 1009
      <212> DNA
      <213> Human

      <400> 107
cgagctatta tggtacggaa cttttttaa tgaggaattt catgatgatt taggaatttt      60
```

```
ctctcttgga aaaggcttcc cctgtgatga aaatgatgtg ccagctaaaa ttgtgtgcca    120
tttaaaaact gaaaatattt taaaattatt tgtctatatt ctaaattgag ctttggatca    180
aactttaggc caggaccagc tcatgcgttc tcattcttcc ttttctcact ctttctctca    240
tcactcacct ctgtattcat tctgttgttt gggatagaaa aatcataaag agccaaccca    300
tctcagaacg ttgtggattg agagagacac tacatgactc caagtatatg agaaaaggac    360
agagctctaa ttgataactc tgtagttcaa aaggaaaaga gtatgcccaa ttctctctac    420
atgacatatt gagatttttt ttaatcaact tttaagatag tgatgttctg ttctaaactg    480
ttctgtttta gtgaaggtag atttttataa aacaagcatg ggattcttt tctaaggtaa     540
tattaatgag aagggaaaaa agtatcttta acagctcttt gttgaagcct gtggtagcmc    600
attatgttta taattgcaca tgtgcacata atctattatg atccaatgca aatacagctc    660
caaaaatatt aaatgtatat atattttaaa atgcctgagg aaatacattt ttcttaataa    720
actgaagagt ctcagtatgg ctattaaaat aattattagc ctcctgttgt gtggctgcaa    780
aacatcacaa agtgaccggt cttgagacct gtgaactgct gccctgttta gtaaataaaa    840
ttaatgcatt tctagagggg gaatatctgc catccagtgg tggaaatgtg gagtaaagaa    900
gctggtggtc tgcttctgtg ctgtatgcca gccttttgcc ttaagttgag aggaggtcaa    960
ctttagctac tgtctttggt ttgagagcca tggcaaaaaa aaaaaaaaa               1009
```

<210> 108
<211> 15
<212> PRT
<213> Homo sapiens

<400> 108
Met Lys Phe Leu Ala Val Leu Val Leu Leu Gly Val Ser Ile Phe
1 5 10 15

<210> 109
<211> 15
<212> PRT
<213> Homo sapien

<400> 109
Gly Val Ser Ile Phe Leu Val Ser Ala Gln Asn Pro Thr Thr Ala
1 5 10 15

<210> 110
<211> 15
<212> PRT
<213> Homo sapien

<400> 110
Asn Pro Thr Thr Ala Ala Pro Ala Asp Thr Tyr Pro Ala Thr Gly
1 5 10 15

<210> 111
<211> 15
<212> PRT
<213> Homo sapien

<400> 111
Tyr Pro Ala Thr Gly Pro Ala Asp Asp Glu Ala Pro Asp Ala Glu
1 5 10 15

<210> 112
<211> 15
<212> PRT
<213> Homo sapien

<400> 112
Ala Pro Asp Ala Glu Thr Thr Ala Ala Ala Thr Thr Ala Thr Thr
1 5 10 15

<210> 113
<211> 15
<212> PRT
<213> Homo sapien

<400> 113
Thr Thr Ala Thr Thr Ala Ala Pro Thr Thr Ala Thr Thr Ala Ala
1 5 10 15

<210> 114
<211> 15
<212> PRT
<213> Homo sapien

<400> 114
Ala Thr Thr Ala Ala Ser Thr Thr Ala Arg Lys Asp Ile Pro Val
1 5 10 15

<210> 115
<211> 15
<212> PRT
<213> Homo sapien

<400> 115
Leu Pro Lys Trp Val Gly Asp Leu Pro Asn Gly Arg Val Cys Pro
1 5 10 15

<210> 116
<211> 15
<212> PRT
<213> Homo sapien

<400> 116
Lys Asp Ile Pro Val Leu Pro Lys Trp Val Gly Asp Leu Pro Asn
1 5 10 15

<210> 117
<211> 621
<212> DNA
<213> Homo sapiens

<400> 117
atgcttcctc ctgccattca tttctatctc cttccccttg catgcatcct aatgaaaagc 60
tgtttggctt ttaaaaatga tgccacagaa atcctttatt cacatgtggt taaacctgtt 120
ccagcacacc ccagcagcaa cagcacgttg aatcaagcca gaaatggagg caggcatttc 180
agtaacactg gactggatcg gaacactcgg gttcaagtgg gttgccggga actgcgttcc 240
accaaataca tctctgatgg ccagtgcacc agcatcagcc ctctgaagga gctggtgtgt 300
gctggcgagt gcttgcccct gccagtgctc cctaactgga ttggaggagg ctatggaaca 360
aagtactgga gcaggaggag ctcccaggag tggcggtgtg tcaatgacaa aacccgtacc 420
cagagaatcc agctgcagtg ccaagatggc agcacacgca cctacaaaat cacagtagtc 480
actgcctgca agtgcaagag gtacacccgg cagcacaacg agtccagtca caactttgag 540
agcatgtcac ctgacaagcc agtccagcat cacagagagc ggaaaagagc cagcaaatcc 600
agcaagcaca gcatgagtta g 621

<210> 118
<211> 618
<212> DNA
<213> Homo sapiens

<400> 118
atgcttcctc ctgccattca tttctatctc cttccccttg catgcatcct aatgaaaagc 60
tgtttggctt ttaaaaatga tgccacagaa atcctttatt cacatgtggt taaacctgtt 120

```
ccagcacacc ccagcagcaa cagcacgttg aatcaagcca gaaatggagg caggcatttc 180
agtaacactg gactggatcg gaacactcgg gttcaagtgg gttgccggga actgcgttcc 240
accaaataca tctctgatgg ccagtgcacc agcatcagcc ctctgaagga gctggtgtgt 300
gctggcgagt gcttgcccct gccagtgctc cctaactgga ttggaggagg ctatggaaca 360
aagtactgga gcaggaggag ctcccaggag tggcggtgtg tcaatgacaa aacccgtacc 420
cagagaatcc agctgcagtg ccaagatggc agcacacgca cctacaaaat cacagtagtc 480
actgcctgca agtgcaagag gtacacccgg cagcacaacg agtccagtca caactttgag 540
agcatgtcac ctgacaagcc agtccagcat cacagagagc ggaaaagagc cagcaaatcc 600
agcaagcaca gcatgagt                                               618
```

<210> 119
<211> 206
<212> PRT
<213> Homo sapiens

<400> 119

```
Met Leu Pro Pro Ala Ile His Phe Tyr Leu Leu Pro Leu Ala Cys Ile
                5                   10                  15

Leu Met Lys Ser Cys Leu Ala Phe Lys Asn Asp Ala Thr Glu Ile Leu
            20                  25                  30

Tyr Ser His Val Val Lys Pro Val Pro Ala His Pro Ser Ser Asn Ser
        35                  40                  45

Thr Leu Asn Gln Ala Arg Asn Gly Gly Arg His Phe Ser Asn Thr Gly
    50                  55                  60

Leu Asp Arg Asn Thr Arg Val Gln Val Gly Cys Arg Glu Leu Arg Ser
65                  70                  75                  80

Thr Lys Tyr Ile Ser Asp Gly Gln Cys Thr Ser Ile Ser Pro Leu Lys
                85                  90                  95

Glu Leu Val Cys Ala Gly Glu Cys Leu Pro Leu Pro Val Leu Pro Asn
            100                 105                 110

Trp Ile Gly Gly Gly Tyr Gly Thr Lys Tyr Trp Ser Arg Arg Ser Ser
        115                 120                 125

Gln Glu Trp Arg Cys Val Asn Asp Lys Thr Arg Thr Gln Arg Ile Gln
    130                 135                 140

Leu Gln Cys Gln Asp Gly Ser Thr Arg Thr Tyr Lys Ile Thr Val Val
145                 150                 155                 160

Thr Ala Cys Lys Cys Lys Arg Tyr Thr Arg Gln His Asn Glu Ser Ser
                165                 170                 175

His Asn Phe Glu Ser Met Ser Pro Asp Lys Pro Val Gln His His Arg
            180                 185                 190

Glu Arg Lys Arg Ala Ser Lys Ser Ser Lys His Ser Met Ser
        195                 200                 205
```

<210> 120
<211> 24
<212> PRT
<213> Homo sapiens

<400> 120

Met Leu Pro Pro Ala Ile His Phe Tyr Leu Leu Pro Leu Ala Cys Ile
5 10 15

Leu Met Lys Ser Cys Leu Ala Phe
20

<210> 121
<211> 182
<212> PRT
<213> Homo sapiens

<400> 121
Lys Asn Asp Ala Thr Glu Ile Leu Tyr Ser His Val Val Lys Pro Val
5 10 15

Pro Ala His Pro Ser Ser Asn Ser Thr Leu Asn Gln Ala Arg Asn Gly
20 25 30

Gly Arg His Phe Ser Asn Thr Gly Leu Asp Arg Asn Thr Arg Val Gln
35 40 45

Val Gly Cys Arg Glu Leu Arg Ser Thr Lys Tyr Ile Ser Asp Gly Gln
50 55 60

Cys Thr Ser Ile Ser Pro Leu Lys Glu Leu Val Cys Ala Gly Glu Cys
65 70 75 80

Leu Pro Leu Pro Val Leu Pro Asn Trp Ile Gly Gly Gly Tyr Gly Thr
85 90 95

Lys Tyr Trp Ser Arg Arg Ser Ser Gln Glu Trp Arg Cys Val Asn Asp
100 105 110

Lys Thr Arg Thr Gln Arg Ile Gln Leu Gln Cys Gln Asp Gly Ser Thr
115 120 125

Arg Thr Tyr Lys Ile Thr Val Val Thr Ala Cys Lys Cys Lys Arg Tyr
130 135 140

Thr Arg Gln His Asn Glu Ser Ser His Asn Phe Glu Ser Met Ser Pro
145 150 155 160

Asp Lys Pro Val Gln His His Arg Glu Arg Lys Arg Ala Ser Lys Ser
165 170 175

Ser Lys His Ser Met Ser
180